Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Stanford University*
Donald B. Lindsley, *Northwestern University*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Margaret K. Harlow, Managing Editor

CONTENTS

Secondary Reinforcement of a Simple Discrimination in Human Beings: W. R. HUBBARD.. 233

Recovery from Retention Loss as a Function of Amount of Pre-recall Warming-up: A. L. IRION AND DOROTHY S. WHAM 242

Facilitation and Interference in Performance on the Modified Mashburn Apparatus: I. The Effects of Varying the Amount of Original Learning: D. LEWIS, DOROTHY E. MCALLISTER, AND J. A. ADAMS 247

Temporal Gradients of Response Strength with Two Levels of Motivation: G. ROSENBAUM 261

The Effects of Differential Rewards on Discrimination Reversal Learning by Monkeys: D. R. MEYER 268

The Effect of Practice with Brief-Exposure Techniques upon Central and Peripheral Visual Acuity and a Search for a Brief Test of Peripheral Acuity: R. H. BRUCE AND F. N. LOW 275

Probability Learning of Perceptual Cues in the Establishment of a Weight Illusion: E. BRUNSWIK AND H. HERMA 281

Probability Learning and a Negative Recency Effect in the Serial Anticipation of Alternative Symbols: M. E. JARVIK 291

The Effect of Scale Numbering on Scale-Reading Accuracy and Speed: J. L. BARBER AND W. R. GARNER 298

The Accuracy of Counting Repeated Short Tones: W. R. GARNER 310

American Psychological Association

Vol. 41, No. 4 April 1951

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.00, or $14.00 annually. Single copies are $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., Prince and Lemon Sts., Lancaster, Pa., or 1515 Massachusetts Avenue, N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental psychology or for behavior theory. Normally, articles of a length exceeding 25 printed pages cannot be accepted. However, a single experimental study or an integrated series of studies accomplished simultaneously must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free reprints without covers. All of the cost of the author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas.

Immediate publication can be given to articles whose authors assume the full cost of publication. Such articles must undergo the usual review by the Consulting Editors of the JOURNAL. No free reprints are given in the case of these articles.

Address all articles submitted for regular publication or for immediate publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

The JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. Anderson and W. L. Valentine, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, 41, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *Es*, *E*'s, and *Es*'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts must be typed double-spaced throughout, including tables, footnotes, references, and titles for figures. Tables, figures, footnotes, references, and titles for figures must be placed on separate pages at the end of the manuscript.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 41, NO. 4 APRIL 1951

SECONDARY REINFORCEMENT OF A SIMPLE DISCRIMINATION IN HUMAN BEINGS[1]

BY WILBUR R. HUBBARD[2]

Indiana University

Secondary reinforcement is a well-established fact of considerable importance in the study of animal behavior. A stimulus or stimulus pattern which acquires reinforcing properties as a consequence of contiguous occurrence with a previously established reinforcing agent, or stimulus pattern, is called a secondary reinforcing agent or secondary reinforcer (**10**). Recently, Spence (**19**) has suggested that learning under conditions of delay of reinforcement occurs through the development of secondary reinforcement in conjunction with stimulus generalization.

Several different aspects of the problem of secondary reinforcement are treated in the more recent studies. There are some which show that a secondary reinforcing agent may delay extinction (**2, 12**), some which indicate that appropriately placed secondary reinforcers either retard or facilitate learning a simple discrimination under conditions of either delay or no-delay of reinforcement (**3, 4, 8, 14**), and others which exhibit generalization of secondary reinforcement (**3, 5, 6**). Still others show the relationship between secondary reinforcement and learning under conditions of partial reinforcement (**2**), and maze learning where the only reinforcement was secondary (**15**). All of these experiments were performed with rats.

It seems to be generally supposed that secondary reinforcement is particularly important in human learning. Most of the reinforcers encountered by human beings presumably have acquired their reinforcing properties in situations similar to those which define secondary reinforcement at the lower animal level. Before the concept of secondary reinforcement can be used widely with respect to human behavior, we need experiments using human subjects and paralleling the work done with infrahuman animals.

STATEMENT OF THE PROBLEM

In 1938, Bugelski (**1**) performed a basic and clear-cut experiment with

[1] A thesis submitted in 1949 to the faculty of the Graduate School of Indiana University in partial fulfillment of the requirements for the degree of Master of Arts. The writer is indebted to Dr. C. J. Burke and particularly to Dr. I. J. Saltzman for guidance in planning and conducting this research. The writer assumes full responsibility for the interpretation of the data.

[2] Now at the University of Oregon.

rats which demonstrated the ability of a secondary reinforcer to delay extinction. Two groups of rats were trained to press a retractable bar in a Skinner Box. Food reinforcements were accompanied by the click of the magazine. The groups were equated on the basis of the number of trials required to obtain 30 reinforcements. One group was given extinction trials with the click of the magazine present, while the other group was treated identically except that the click was absent. The criterion of extinction was 5 min. of no responding. The "click group" gave 30 per cent more responses than the "no click group." There was no control for the effects of the click per se.

The purpose of the present experiment was to see what results would be obtained with human *S*s subjected to conditions comparable to those which would eventuate in secondary reinforcement in subhuman *S*s. Bugelski's design was used, the response being a simple discrimination. A third group was added to control for the effects of the "secondary reinforcing stimulus" not attributable to its being paired with the reinforcement.[3]

Green light:	ON	OFF									ON	OFF	
Shutter:		Open							Closed			Open	
Time in seconds:	0	1	2	3	4	5	6	7	8	9	10	11	12

Apparatus and Procedure

Subjects.—Eighty-one undergraduate men and women from psychology and journalism classes in Indiana University were randomly assigned to three groups, so that of each three *S*s one was assigned to each group. Postexperimental interrogations revealed that all *S*s were naive insofar as this experiment was concerned.

Apparatus.—The apparatus and most of its dimensions are represented in Fig. 1. The vertical panel separated *S* and *E*, and concealed the working parts of the apparatus from *S*. The ¾-in. by 1-in. aperture was 2 in. from the top of the panel. The green pilot lamp was 5½ in. to the left of the aperture, while the red one was 5½ in. to the right. Both lamps were 1 in. in diameter. The keys were mounted in a 1-in. deep recess so that their tops were flush with the top of the table. They were spring-loaded with a 1-in. throw, and placed on an arc so that the one on the left was closest to *S*. This arrangement allowed *S* to reach any key with his index finger without moving his arm. In Fig. 1, *S*'s side of the apparatus is in the foreground. A shutter was mounted just behind the aperture. When the shutter was in the open position, the number *1*, *2*, or *3* could be seen. Each key corresponded to one of the numbers, but did not bear its number. The order of the keys from left to right was: *3*, *1*, *2*. The numbers were ¾ in. wide and ¼ in. high, of plain block design, and appeared in random order in such a manner that each number occurred once in every group of three.

The apparatus was automatic in operation. When *E* turned on the master switch, the green lamp lighted for 1 sec., then the shutter opened, exposing a number for 8 sec., and finally, the shutter closed for 2 sec. One second after the shutter closed the green lamp lighted again. The apparatus continued this sequence until the master switch was turned off. A summary of the timing and sequence of events follows:

If, while the shutter was open, *S* pressed the key corresponding to the number in the aperture, the red light flashed for 0.5 sec. If any other key was pressed, the red light did not flash regardless of whether or not *S* subsequently made the correct choice during that trial. Moreover, the red light flashed only once during any shutter-open period. The green light was designated as the signal that the shutter was going to open. The red light was designated as the sign that the correct choice had been made.

A Hewlitt-Packard audio oscillator was set at 850 c.p.s. and 61 db. above the standard threshold.[4] The speaker circuit was so wired that the tone sounded *contiguously with*, or *in place of*, the red light. Whether or not the apparatus presented the red light, the tone, or both was

[3] At this point, the term "secondary reinforcing stimulus" is used only to designate the stimulus that merits the term later.

[4] The intensity of the tone was measured with a General Radio sound level meter placed in the space usually occupied by *S*'s head. The figure given is the mean of a distribution of 40 readings ranging from 58 to 65 db.

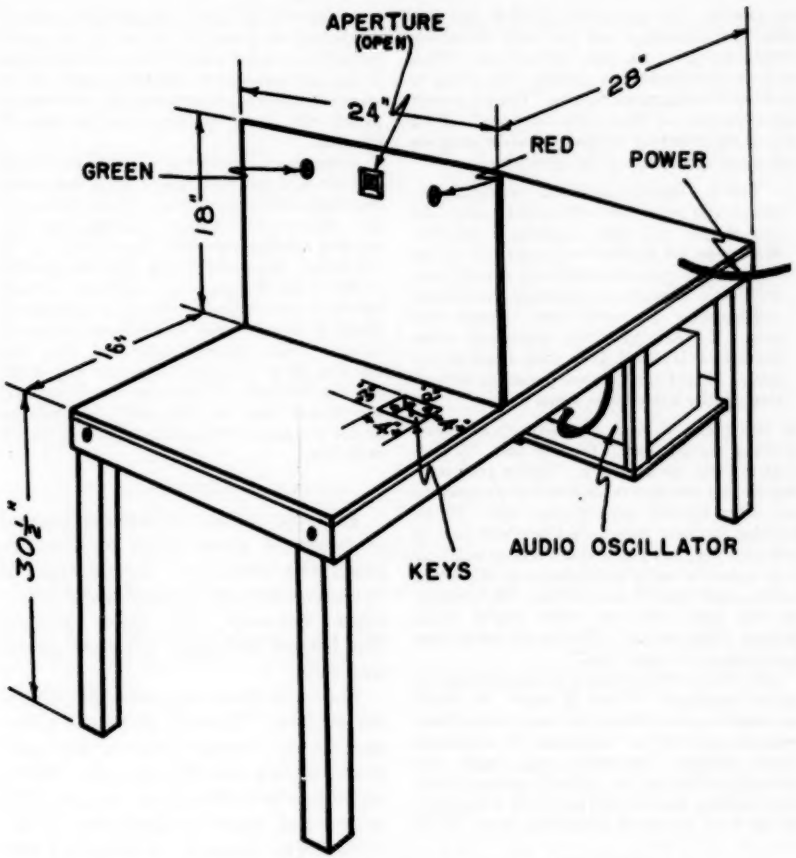

FIG. 1. The apparatus. The *S* was seated in the foreground while *E* was seated at the other end.

determined by the silent operation of two knife switches on *E*'s side of the panel. The *S* was presented with the problem of discovering which key corresponded to which number. The time from one green light to the next was a trial, and the depression of the key corresponding to the number was the correct choice.

On *E*'s side of the panel, a .01 sec. chronoscope was wired to start as soon as the shutter opened and stop at the slightest touch on any one of the keys. A small pilot light, not visible to *S*, flashed whenever *S*'s first choice was correct. This permitted recording the number of correct choices as well as measuring response latencies. The numbers which appeared through the aperture were drawn on a 10-in. circular card turned by a stepping relay. All possible combinations of *1*, *2*, and *3* were randomly assigned to positions on the card. Since the stepping relay had 40 positions, it was necessary to add one more number (*2*) to the card in order to maintain a constant period of time between numbers. Each *S* received the same sequence of numbers. No sounds, lights, or movements of *E* were peculiar to the operation of the red light and tone. A fan served to mask certain barely audible sounds that were not completely eliminated by insulating their sources.

Procedure.—All subjects were run in a sound-insulated room where *E* was the only other per-

son present. He said nothing to *S* at any time after the apparatus was put into operation. Each *S* was given the same instructions. First, in a conversational tone *E* said, "I'm going to read some instructions to you. I can't answer any questions, so listen very carefully." Then *E* read the following instructions while pointing out appropriate parts of the apparatus:

> This is a learning problem. At regular intervals this green light will come on, and then this shutter will open exposing a number. When you see the number, press one of the keys. If you press the right one, this red light will flash. If you press the wrong one, nothing will happen. You won't have to press very hard. Use only your first finger, and when the shutter is closed, keep your finger off the keys. Since I can't answer questions, let's go through the instructions again.

At this point, *E* read the instructions again, omitting the last line. Then he said, "It's all right to press the keys now. Adjust your position so that you can reach any key as easily as any other without moving your arm. I'll be listening for some things, so I won't be able to talk with you, and please don't talk to me." *E* then donned a set of headphones to discourage talking, and took his seat, saying, "Remember, the red light tells you when you're right. Ready? Here we go." The master switch was turned on at the word "go."

The criterion of learning was nine consecutive correct responses. When *S* made the ninth consecutive correct choice, he was given further practice until he had completed 21 additional correct choices. Extinction trials began immediately following the thirtieth correct choice after learning, and without any kind of warning. All *S*s were given 40 extinction trials. If *S* was suspected of having enough knowledge of psychology to guess the nature of the experiment, he was questioned at the end of the session. It was not necessary to eliminate any *S* on this account; however, experimentation was discontinued with three *S*s who failed to learn in 60 trials.

Experimental conditions.—Group I was trained with the light and tone present, and given extinction trials with both absent. When the correct key was pressed during training, the tone sounded contiguously with the flashing of the red light. When the wrong key was pressed, neither *E* nor the apparatus made any response. Group II was trained in the same situation as Group I, but the tone was present during extinction trials. During extinction, the tone sounded when the correct response was made, but the red light did not operate. Group III was trained with the light only, and was accorded the same treatment as Group II during extinction.

TABLE I

COMPARISONS OF MEAN CORRECT CHOICES DURING FORTY EXTINCTION TRIALS

Groups Compared	Mean	Difference	d*	P
I	15.00	18.19	12.167	<.01
vs.				
II	33.19			
I	15.00	11.04	8.315	<.01
vs.				
III	26.04			
II	33.19	7.15	6.055	.01
vs.				
III	26.04			

* d is the statistic computed in the Festinger test.

RESULTS

The rate of learning was determined in terms of mean trials required to reach the criterion. An analysis of variance revealed no significant differences between the three groups. The largest difference between means was 1.29.

The extinction data present a different picture. Table I gives the mean number of correct choices for each group during the 40 extinction trials, as well as the differences between the means and their probabilities of occurrence by chance. Festinger's test was employed because it was not possible to meet the assumptions required for either the t-test or analysis of variance.[5] The differences between

[5] The variances of the distributions of correct choices were clearly not equal, and the distributions for Groups II and III were markedly skewed and nearly rectangular, respectively. The use of χ^2 was precluded by the lack of independence of the individual measures. The Festinger test is based on the ranks of pooled samples. The hypothesis tested is that the two samples are drawn from the same population, but the test is believed to be most sensitive to differences between means. The test requires independent random samples. The steps in computation are listed in the article cited.

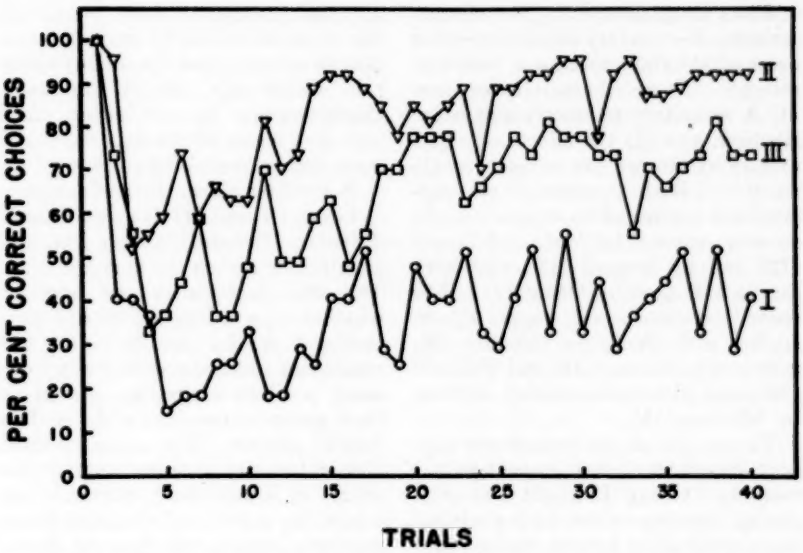

FIG. 2. The per cent correct choices as a function of the number of extinction trials

the means were significant at the 1 per cent level of confidence or better in every instance.

By examining Fig. 2, where the per cent correct choices is plotted as a function of the number of extinction trials, it can be seen that all groups show an early and rapid decrease in the percentage of correct choices, followed by a gradual increase and leveling off. The per cent correct choices expected to occur by chance is 33⅓. Groups I and III approximate chance expectancy at trial 4, whereas Group II never reached the chance level. Following trial 4, the performance of Group I drops to about 15 per cent. Group III shows no such decrease at any time. The performance of Group II is rather consistently higher than that of Group III, which is in turn higher than that of Group I. The curves reach their minima in the order II, III, I, but since no psychologically meaningful function was found which would fit all three, it was not possible to test the null hypothesis with regard to either the rate of extinction or the acceleration. The curve for Group I appears to level off at about 40 per cent, for Group II at about 91 per cent, and for Group III at about 73 per cent. Whether or not the curves actually approach asymptotes is not known; curve-fitting operations were unsuccessful again. Response-latency measures were so variable as to be of no use. The latency scores ranged from 0.01 second to 6.14 seconds, and their means showed no consistent trends.

DISCUSSION

The purpose of this study was to subject human beings to operations comparable to those which eventuate in the development of secondary reinforcement in subhuman animals, and to note the consequent similarities and differences in animal and human behavior in such a situation.

From Bugelski's study two characteristics of secondary reinforcement in cases of infrahuman animal behavior emerge. These characteristics are: (1) A secondary reinforcer delays extinction; and (2) the effects of a secondary reinforcer are subject to extinction. Both statements are supported in studies with rats in a straight runway reported by Mote and Finger (**12**) and by Seward (**17**), and with rats in a T-maze by Denny (**2**). The second statement is supported in studies with chicks by Grindley (**9**), with rats by Skinner (**18**) and Williams (**20**), and with feebleminded children by Mitrano (**11**).

The results of this experiment support statement 1 with respect to human *Ss*. Group II (light and tone during training—tone during extinction) made more correct choices during the extinction trials than Group I (light and tone during training—neither during extinction). Extinction may be defined in this situation as *the occurrence of correct choices with a frequency expected by chance*, or 33⅓ per cent. Extinction occurred in Group I, and possibly in Group III, but almost certainly not in Group II. Group III (light during training—tone during extinction) was introduced to control for effects of the tone that might occur without its being paired with the red light. Table I and Fig. 2 reveal that the performance of Group III was consistently below that of Group II and above that of Group I, differing significantly from both of them. Thus, we may say that a secondary reinforcing agent delays or prevents the extinction of a simple discrimination by human *Ss*. However, the results of this study do not indicate that the effects of the secondary reinforcing agent are subject to extinction in human *Ss*. The rises in the extinction curves preclude any such statement at this time. At the same time, it is by no means certain that rats trained in a discrimination would not exhibit the same characteristic. In any event, there are two aspects of the data that raise some rather interesting questions.

In the first place, the performances of Groups I and III are significantly different. Secondly, each curve rises after falling during the first few trials. The first fact raises the question whether or not Bugelski would have obtained similar results if he had trained an additional group with food, using a silent magazine, and given that group extinction trials with a "click" present. The second of these facts raises the question whether the effects of a secondary reinforcer are subject to extinction in human *Ss*, or whether, instead, the rises are due to some artifact of the technique used in this experiment. In addition to these two questions there are still others which are bound to arise from whatever answers are given; such questions as:

a) If the results from Group III cannot be duplicated with infrahuman animals, why not?

b) If the rises in the extinction curves are not due to an artifact, but are instead due to a genuine characteristic of secondary reinforcement in cases of discrimination, to what may such a characteristic be attributed?

An explanation of the behavior of Group III can be made in terms of generalization. Denny (**3**) found generalization of secondary reinforcement with rats, and so did Saltzman (**16**). Considering their results, and noting that human *Ss* are repeatedly reinforced in everyday life for behavior involving generalization, we believe it not improbable that we have

a case of cross-modality generalization of reinforcement in Group III. Returning to a consideration of the procedures in this study, we find several circumstances that favor such an explanation. First, the red light and tone have at least three characteristics in common: (1) duration of stimulation, (2) sudden onset and cessation, and (3) occurrence dependent on the depression of a particular key for a particular number. Secondly, nothing in the situation changed except for the cessation of the operation of the red light and the beginning of the operation of the tone. There was no break in the functioning of the apparatus, no signal from *E*, and no sound from the manipulation of the switches. These circumstances are admirably suited to the occurrence of generalization. Hence, the prediction on these bases would be that Group III would make more correct choices than Group I because of the generalization of reinforcement from the red light to the tone in the case of Group III. Further, since most studies of generalization show that generalized response strength is less than the response strength in the original conditioning situation, we would predict that Group III would show considerably less than perfect performance. That these predictions are borne out by the results of this experiment can be seen readily in Table I and Fig. 2.

If the generalization hypothesis is correct, results similar to these might be found with rats, provided that "primary" and "secondary" reinforcers are found which are similar enough to allow for a significant amount of generalization. As pointed out previously, the red light and tone were similar in several respects.

There are three sets of circumstances to be considered in handling the problem of whether or not the shape of the extinction curves (Fig. 2) is due to an artifact of the experimental procedure. First, it is nearly inconceivable that the *S*s failed to engage in some implicit verbal reactions during the experiment. The data provide no basis for estimating the nature of the role played by such reactions. A second consideration is that human *S*s are repeatedly reinforced in everyday life for rapidly abandoning nonrewarded discriminations and forming others in their stead. Thirdly, the *S*s in this experiment frequently made comments concerning the experimental procedure. The frequency of such comments as, "You changed the pattern," "Hey, the red light doesn't work," "The little horn was it, wasn't it?" and "I couldn't tell what was right for a while" suggest that the extinction periods became new learning situations—not in the sense that the *S*s were learning not to make the correct choice, but in the sense that they were learning that the original discrimination was correct, and that only the indicator of that fact had changed.

The writer has no hypotheses as to just what variables might be involved in an analytic explanation of this phenomenon. It would be expected that depression of the different keys would provide different sets of stimulation which would acquire reinforcing properties as did the tone; thus, Group I could be expected to show some increase in performance for reasons similar to those that lead to the greater increases in Groups II and III. Another explanation of the rise exhibited by Group I can be made in terms of Hull's reactive inhibition. At the beginning of the extinction period, when nonrewarded correct choices are made frequently, reactive inhibition builds up rapidly, resulting

in a decreased frequency of correct choices. As the frequency decreases, more reactive inhibition is dissipated between correct choices, which results in a gradual increase in the frequency. The per cent correct choices approaches a stable level where the rate of increase in reactive inhibition is equal to its rate of dissipation. If this explanation is applied to the performance of Group I, it must also be applied in the other two cases, and the validity of such an application is clearly doubtful. Therefore, our tentative answer to question 2 is that the rises in the extinction curves may well be an artifact of the experimental technique. If so, human *Ss* should behave more like infrahuman *Ss* if the required reaction was simply the depression of one key for some sort of reinforcement. However, an additional check is needed, a test of the hypothesis that the rises of the extinction curves are a genuine characteristic of secondary reinforcement in cases of discrimination. A Skinner-type box with three bars and corresponding discriminable pilot lights might serve the purpose. In either of these two cases, the additional condition that the original and secondary reinforcing stimuli be markedly different, would provide a test of the hypothesis that the performance of Group III was due to generalization.

Summary and Conclusions

Two groups of college students were trained in a simple discrimination situation in which a tone was paired with a verbally determined reinforcer. One group was given extinction trials with the tone present, and another group with the tone absent. The group for which the tone was present made more correct choices during extinction than did the other. A third group, trained with only the original reinforcer and given extinction trials with the tone present, gave a performance that fell between those of the first two groups. During the extinction period, the performance curves of all three groups rapidly reached their minima and subsequently increased and leveled off. The differences between group means were significant at the 1 per cent level of confidence. The foregoing data seem to warrant the following conclusions:

1. Secondary reinforcement may serve to delay the extinction of a response in human beings.

2. A stimulus not previously paired with a reinforcing stimulus may behave like a secondary reinforcer, but is not as effective.

3. These results suggest that secondary reinforcement operates in human behavior in much the same manner that it does in the behavior of infrahuman animals.

(Manuscript received March 14, 1950)

References

1. Bugelski, R. Extinction with and without sub-goal reinforcement. *J. comp. Psychol.*, 1938, **26**, 121–133.
2. Denny, M. R. The role of secondary reinforcement in a partial reinforcement learning situation. *J. exp. Psychol.*, 1946, **36**, 373–389.
3. Denny, M. R. The effect of using differential end boxes in a simple T-maze learning situation. *J. exp. Psychol.*, 1948, **38**, 245–250.
4. Ehrenfreund, D. Effect of a secondary reinforcing agent in black-white discrimination. *J. comp. physiol. Psychol.*, 1949, **42**, 1–6.
5. Estes, W. K. A study of motivating conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1949, **39**, 306–311.
6. Estes, W. K. Generalization of secondary reinforcement from the primary drive. *J. comp. physiol. Psychol.*, 1949, **42**, 286–295.

7. Festinger, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika*, 1946, **11**, 71–81.
8. Grice, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, **38**, 1–16.
9. Grindley, G. C. Experiments on the influence of the amount of reward on learning in young chickens. *Brit. J. Psychol.*, 1929, **20**, 173–180.
10. Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
11. Mitrano, A. J. Principles of conditioning in human goal behavior. *Psychol. Monogr.*, 1939, **51**, No. 230.
12. Mote, F. A., and Finger, F. W. Exploratory drive and secondary reinforcement in the acquisition and extinction of a simple running response. *J. exp. Psychol.*, 1942, **31**, 57–68.
13. Nissen, H. W., and Crawford, M. P. A preliminary study of food sharing behavior in young chimpanzees. *J. comp. Psychol.*, 1936, **22**, 383–419.
14. Perkins, C. C., Jr. The relation of secondary reward to gradients of reinforcement. *J. exp. Psychol.*, 1947, **37**, 377–392.
15. Saltzman, I. J. Maze learning in the absence of primary reinforcement: A study of the effective duration of secondary reinforcing properties. *J. comp. physiol. Psychol.*, 1949, **42**, 161–173.
16. Saltzman, I. J. Generalization of secondary reinforcement. *J. exp. Psychol.*, 1950, **40**, 189–193.
17. Seward, J. P. Expectancy in the extinction of a conditioned operant response. *Amer. Psychologist*, 1948, **3**, 242. (Abstract.)
18. Skinner, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.
19. Spence, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, **54**, 1–8.
20. Williams, K. A. The reward value of a conditioned stimulus. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 31–55.

RECOVERY FROM RETENTION LOSS AS A FUNCTION OF AMOUNT OF PRE-RECALL WARMING-UP

ARTHUR L. IRION AND DOROTHY S. WHAM

University of Illinois

In a recent article, Irion (**4**) has emphasized the importance of *S*'s performance set as a condition determining amount of forgetting in verbal learning. According to this view, forgetting will occur if *S* changes his performance set during the retention interval from one which relatively facilitates the performance of the originally learned activity to one which is, in this respect, less adequate. Under such circumstances, recall performances should be depressed, and it is to be anticipated that the relearning performance curve will show a sharp initial rise. This swift improvement during the initial stages of relearning has been termed the warming-up effect and may be identified with that period during relearning in which *S* regains a performance set which facilitates performance of the originally learned activity.

On the other hand, maintenance of an adequate performance set throughout the retention interval should result in a reduced decrement in recall. Furthermore, reinstatement of an adequate performance set, as by giving *S* a number of warming-up trials on some similar task immediately before recall, should be followed by a decrease in forgetting. Under such circumstances, recall performance should show a reduced retention loss, and there should be a corresponding reduction in the rate of increase in proficiency noted during the initial stages of relearning.

Previously obtained experimental results (**5**) indicate that if *S*s enter relearning after engaging in a warming-up activity immediately before recall, the amount of forgetting which normally follows the introduction of a rest period is sharply decreased. Presumably, the pre-recall warming-up activity served to reinstate in the *S*s a performance set which facilitated performance of the learned activity. Such re-establishment of performance set by the technique of giving warming-up trials is, of course, dependent upon existence of a warming-up task which is sufficiently similar to the activity being learned for the warming-up to be effective without, at the same time, affording practice to the *S*s in the specific stimulus-response relationships which have been acquired during original learning. Furthermore, in order to avoid possible transfer-of-training effects, it is desirable that little learning take place as a result of engaging in the warming-up activity.

The present study is concerned with determination of the relationship between amount of pre-recall warming-up and subsequent recall and relearning performance and is, thus, a test of the following theoretical prediction made by Irion (**4**):

> Regardless of the activities which occur during the rest period, if, immediately prior to measuring retention, the subject engages in an activity which reinstates an appropriate set to perform, retention should be greater than if this set had not been reinstated. In this case, the extent of recovery from retention loss should be a function of the similarity between the set-inducing task and the learning task, *and of the amount of time devoted to the induction of appropriate set.*[1]

[1] Italics added.

In the retention of nonsense syllables partially learned by the rote–serial anticipation method, varying amounts of warming-up activity were introduced immediately before recall and relearning. Specific predictions of experimental results can be made, as follows:

1. The greater the amount of pre-recall warming-up activity, the higher should be the level of recall performance.

2. The greater the amount of pre-recall activity, the less rapidly should proficiency increase during relearning.

Procedure

The learning task employed in this experiment was the memorization of a list of nine nonsense syllables by the rote–serial anticipation method. The *Ss* were required to spell, rather than to pronounce, the syllables. The syllables making up the list were taken from Glaze's list of syllables (from those having 47 and 53 per cent association value) as reported by Melton (6), and the list was prepared according to the rules suggested by Melton (6).

This list of nine nonsense syllables, together with the symbol ••• which preceded the first syllable, was presented to *Ss* on a Missouri-type memory drum at the rate of 2.18 sec. per syllable. The trials were separated by rests of 4.36 sec., measured from the disappearance of the last syllable in the list to the appearance of the symbol which preceded the first syllable. The memory drum was housed in a screen containing an aperture through which the materials to be learned could be seen by *S*. This screen served effectively to conceal the apparatus and *E* from *S*, and, in general, to reduce the number of distracting stimuli.

Ninety *Ss* (67 males and 23 females) were divided at random into six experimental groups of 15 *Ss* each. All *Ss* were volunteers from courses in general psychology at the University of Illinois. No *S* served in more than one condition of the experiment, and all were naive with respect to this type of experimental situation. No pre-training was given to any *S*.[2]

[2] This departure from more conventional procedure was motivated by belief that the amount of warming-up effect is a function of the total amount of previous practice and the number of times the *S* has previously warmed up on a similar activity.

The experimental design demanded the interpolation of rest intervals between original learning and relearning in several of the experimental conditions. These rest intervals were filled with cartoon-reading. It is thought that this activity is not conducive to rehearsal of the learned material, nor does it cause any great amount of new learning to occur. The same cartoons were available to all *Ss* under all conditions wherein rest intervals were introduced, and *Ss* were encouraged to read these materials during the entire rest interval. All questions concerning the nature of the experiment were deferred until the experimental work had been completed.

The warming-up activity which was used in this experiment was, of necessity, one which would be highly similar to the rote–serial anticipation learning of the list of nonsense syllables, yet one which would not afford specific practice on any portion of that list. Furthermore, it was considered desirable that little learning should occur during the warming-up period in order that possible transfer-of-training effects might be avoided. The warming-up task which was selected was the recitation of a list of nine three-place numbers. A list of these numbers was prepared together with a preceding symbol, after the fashion of the list of nonsense syllables. This list was presented on the memory drum, and the *Ss* were required to anticipate the three-place numbers. In order to minimize the possibility of learning taking place during the warming-up period, the three-place numbers were arranged in order, the first being 123, the second, 456, the third, 789, the fourth, 123, and so on. Since *Ss* were informed of the arrangement of the numbers before the warming-up period, and since each *S* was proficient in counting from 1 through 9, it was possible for all the *Ss* to anticipate correctly each number on the first warming-up trial. It is felt that the possibility of further learning occurring during the warming-up period was extremely slight.

The conditions of this experiment are outlined in Table I. It will be noted that, with the ex-

TABLE I

Plan of Experiment

Condition	No. Trials Original Learning	Length of Rest	No. Trials Warming Up	No. Trials Relearning
I	20	no rest	0	10
II	20	35′0″	0	10
III	20	34′50″	0.5	10
IV	20	34′36″	1	10
V	20	34′12″	2	10
VI	20	33′24″	4	10

ception of Condition I, all conditions were run under a constant procedure except that amount of pre-recall warming-up activity was varied. The no-rest condition (Condition I) was included in order that a determination of amount of forgetting might be made, and in order that, should warming-up prove exceptionally efficacious in reducing retention loss, a control condition for the reminiscence phenomenon should exist.

Results

Original learning.—The means for the last trial of original learning (trial 20) and the first trial of relearning (trial 21) are presented in Table II. In order to determine whether or not significant differences in original learning existed between the six experimental groups of this study, two analyses of variance were conducted, one for the number of correct anticipations on the last original learning trial and the other for the total number of correct anticipations during original learning. In neither case did the obtained differences approach statistical significance.

Recall and relearning.—The mean number of correct anticipations on the first relearning trial (trial 21) is shown in Table II for each experimental condition. It will be seen that all the conditions under which *S*s had some amount of warming-up before relearning were superior in recall to Condition II wherein *S*s received no warming-up before recall. The scores on the first relearning trial (under Conditions II through VI) were subjected to an analysis of covariance, the total number of correct anticipations during original learning being used as initial measures. This analysis resulted in an F of 5.03 for 5 and 69 degrees of freedom, a value which is significant beyond the 1 per cent level of confidence. It may be concluded, then, that amount of retention varies as some function of degree of pre-recall warming-up.

When recall scores are considered alone, the functional relationship between amount of warming-up and degree of retention is somewhat obscured by compensatory differences in original learning proficiency. When these differences are eliminated, as in considering the change in proficiency between the last original learning trial and the first relearning trial, this relationship is more easily noted. The last column in Table II shows the dependency of recall scores on the amount of pre-recall warming-up in terms of the change in the mean number of correct anticipations between trial 20 (the last original learning trial) and trial 21 (the first relearning trial).

In Table II it will be noted that, for *S*s in Condition VI (the condition affording the most warming-up activity prior to relearning), improvement occurred over the 35-min. retention interval. This improvement cannot, however, be identified with the reminiscence phenomenon except by a comparison with a comparable measure taken from Condition I (the condition which did not offer a rest period between trials 20 and 21), since the amount of improvement to be expected from a single additional trial of practice would not otherwise be controlled. It will be further noted that, as between

TABLE II

Mean Correct Anticipations on Trials Twenty and Twenty-one

Condition	Warming-Up Trials	Mean Correct Anticipations		Difference
		Trial 20	Trial 21	
I	—	6.00	6.00	0.00
II	0	6.07	4.40	−1.62
III	0.5	6.00	5.67	−0.33
IV	1	6.20	5.46	−0.74
V	2	5.87	5.53	−0.34
VI	4	4.53	5.87	+1.34

Conditions I and VI, a mean difference of 1.34 correct anticipations (measured in terms of improvement between trials 20 and 21) exists in favor of Condition VI. The t for this difference is 2.20 for 28 degrees of freedom, a value which is significant at between the 5 and 2 per cent levels of confidence. This suggests that reminiscence can occur in the rote–serial anticipation learning of nonsense syllables well beyond the limits of the 5-min. retention interval to which its occurrence has heretofore been limited, providing that the *S*s enter relearning in a warmed-up condition.[3]

In the introductory section of this paper, it was noted that, if *S* entered relearning after having been warmed up on some similar task, the rate of rise of the relearning curve should not be as great as if this warming-up had not been administered prior to relearning. When the change in proficiency between the first two relearning trials is taken as an indication of the initial slope of the relearning function, the relationship between slope of the relearning curve and amount of warming-up activity administered prior to relearning is shown in Table III. There is, it seems, a general tendency for the slope of the relearning curve to be steeper under conditions wherein little or no warming-up was given prior to relearning. An analysis of variance of these differences resulted in an F of 2.62 for 4 and 70 degrees of freedom, a value which is significant at the 5 per cent level of confidence, thus tending to substantiate the theoretical prediction.

TABLE III

MEAN INCREASE IN CORRECT ANTICIPATIONS BETWEEN THE FIRST AND SECOND RELEARNING TRIALS

Condition	Description	Mean Increase
II	No warming-up trials	1.53
III	0.5 warming-up trial	0.93
IV	1 warming-up trial	0.60
V	2 warming-up trials	0.87
VI	4 warming-up trials	0.60
I	No rest—maximum warming-up	0.13

SUMMARY

1. The effects upon retention of pre-recall warming-up were demonstrated in the rote–serial anticipation learning situation. Previous investigation of these effects had been confined to the paired-associate learning situation.

2. Amount of recovery from retention loss was shown to be some increasing function of amount of pre-recall warming-up.

3. Rate of rise of the relearning function tended to be greater following small than following large amounts of pre-recall warming-up activity.

4. There was a suggestion that reminiscence might occur over a 35-min. retention interval under conditions of maximum warming-up before recall. In view of the previously established temporal limitations of the reminiscence phenomenon, a conservative view of this possibility is suggested until such time as confirmatory evidence is obtained.

(Manuscript received March 20, 1950)

[3] See Ward (7) and Hovland (1, 2, 3). A slight difference in methodology exists between the studies cited above and the present study. Ward and Hovland carried out original learning to a performance criterion rather than to a work criterion. This tended greatly to reduce differences in final original learning proficiency. In the present study, differences in original learning proficiency are discounted by using the increase in proficiency between the last original learning trial and the first relearning trial in determining reminiscence, rather than by learning to a performance criterion and then making a direct comparison between conditions on the first relearning trial.

References

1. Hovland, C. I. Experimental studies in rote-learning theory: I. Reminiscence following learning by massed and distributed practice. *J. exp. Psychol.*, 1938, **22**, 201–224.
2. Hovland, C. I. Experimental studies in rote-learning theory: II. Reminiscence with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **22**, 338–353.
3. Hovland, C. I. Experimental studies in rote-learning theory: IV. Comparison of reminiscence in serial and paired-associate learning. *J. exp. Psychol.*, 1939, **24**, 466–484.
4. Irion, A. L. The relation of 'set' to retention. *Psychol. Rev.*, 1948, **55**, 336–341.
5. Irion, A. L. Retention and warming-up effects in paired-associate learning. *J. exp. Psychol.*, 1949, **39**, 669–675.
6. Melton, A. W. Materials for use in experimental studies of the learning and retention of verbal habits. (Mimeographed manuscript.)
7. Ward, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, **49**, No. 220.

FACILITATION AND INTERFERENCE IN PERFORMANCE ON THE MODIFIED MASHBURN APPARATUS: I. THE EFFECTS OF VARYING THE AMOUNT OF ORIGINAL LEARNING[1]

BY DON LEWIS, DOROTHY E. McALLISTER, AND JACK A. ADAMS[2]

State University of Iowa

The problem was to determine the kinds and amounts of facilitation and interference in psychomotor performance resulting from three different amounts of practice on an original task and four different amounts of practice on an interpolated task. The tasks were predominantly motor in character and were available on a single device—the Iowa version of the Mashburn apparatus.

This report is concerned primarily with the effects of varying the amount of OL (original learning). A second report (8) is concerned with the effects of varying the amount of IL (interpolated learning). The amounts of OL and IL were determined by the number of practice trials.

There have been no previous investigations of facilitation and interference in motor learning which have any direct bearing on the problem. In fact, prior to 1947, there had been no satisfactory basis for making a systematic study of facilitative and interfering effects. Experiments in the Iowa Laboratory (4), begun in 1941 but soon interrupted by the war and not renewed until 1947, were the first to utilize motor tasks with which unmistakable signs of facilitation and interference could be obtained. Previous experiments, including those of Buxton and Henry (1) and Buxton and Grant (2), had not utilized tasks in OL and IL which induced either proactive or retroactive effects in clearly recognizable forms.

Interference in *verbal learning* has been studied by numerous investigators, and McGeoch (9) has specifically studied the effects on retroactive inhibition of different levels of OL, obtaining results for relative amounts of retroaction which differed from those for absolute amounts. The findings on verbal learning provide no firm basis for predicting anything about facilitation and interference in motor learning.

Apparatus

Use was made of the modified Mashburn apparatus which has been described by Lewis and Shephard (5). The apparatus is basically similar to the instrument designed in the early 1930's by Mashburn (10), and also to the SAM Complex Coordination Test (11). It consists of three double banks of small pilot lights and a control stick and rudder bar. Each double bank has a row of 13 red stimulus lights and a parallel row of 13 green response lights. Various combinations of stimulus lights come on automatically in random order to provide patterns of red lights that must be matched by corresponding green lights. The three rows of green response lights are separately controlled by means of commutators and brushes associated with the stick and rudder bar. The green lights in any row come on and go off in succession as the corresponding control is moved in a given direction.

The *S* is presented with three randomly selected red lights, one in each of the three double banks. His task is to manipulate the stick and rudder until a green light appears opposite each of the three red lights. When this

[1] This work was done under Contracts N5ori-57 and N9onr-93801 between the State University of Iowa and the Special Devices Center, Office of Naval Research.

[2] Alfred H. Shephard, Walter Spieth, and James L. Hedlund gave valuable assistance in running the *S*s and analyzing the data.

occurs, a stepping relay brings up a new combination of red lights. This matching procedure may be continued for any desired length of time; and *S*'s attainment is indicated by the number of three-way matches accomplished.

A special feature of the Iowa model of the Mashburn is a provision for reversing the direction of movement of each control. The *standard* arrangement of the controls gives a maximum amount of eye-hand and eye-foot coordination. The green lights in the upper horizontal bank (aileron) come on in order from left to right as the stick is moved from left to right; in the vertical bank (elevator) they come on successively from above downward as the stick is pushed forward; and in the bottom horizontal row (rudder) they come on from left to right as the rudder bar is pushed with the right foot. When the controls are *reversed*, the movements of the controls are exactly opposite to those for the standard arrangement, but the movement of the green lights remains the same.

In the present experiment, the *standard task* was always used for OL (and for RL), while the *reversed task* was always used for IL.

A second feature of the modified Mashburn is the provision for recording incorrect responses (false moves) made by *S* immediately after each new set of stimulus lights is presented. When a new set of red lights appears, *S* can move each control either correctly or incorrectly—in the direction appropriate to achieving a match of the new stimulus light, or in the inappropriate direction. A movement of a control in the inappropriate direction is counted as an error. *Three errors are possible on each match.* Separate error counters are provided for each control, thus permitting separate or combined analyses.

TABLE I

THE EXPERIMENTAL GROUPS

Group*	*N*	Amount of OL		Amount of IL		Amount of RL	
		Days	Trials	Days	Trials	Days	Trials
15	8	1	10				
18	10	3	30	1	10		
21	10	5	50				
30	10	1	10				
29	10	3	30	2	20		
31	10	5	50				
16	10	1	10			2	20
19	8	3	30	3	30		
22	9	5	50				
17	10	1	10				
20	10	3	30	5	50		
3	11	5	50				

* The group numbers have no significance, except for use by *E* in identifying groups in this and related studies.

TABLE II

CONTROL GROUPS

Group	*N*	OL		No Practice	IL		RL	
		Days	Trials	Days	Days	Trials	Days	Trials
I	9	5	50	5	0	0	2	20
II	9	10	100	0	0	0		
III	9	3	30	5	0	0		
5	15	0	0	0	2	20	0	0

EXPERIMENTAL DESIGN

Several exploratory studies provided the basis for designing the experiment. Practice trials lasting 2 min. were found to be suitable. A pause of 15 sec. between trials was required to enable *E* to record scores. Blocks of ten trials could be given during each daily session, with a 2-min. rest between trials 5 and 6, without any apparent decrease in the interest or motivation of the *S*s.

The exploratory studies showed that use should be made of experimental conditions providing for variations in amount of OL from around 10 to around 50 trials, with similar variations in the number of IL trials, in order to span the range from small to very large amounts of retroactive interference.

The design involved 12 experimental groups, as listed in Table I. Four amounts of IL were specified: 1, 2, 3, and 5 days, or 10, 20, 30, and 50 trials. For each amount of IL, there were three amounts of OL: 1, 3, and 5 days, or 10, 30, and 50 trials. All 12 groups relearned the original task on two days (20 trials).

Four control groups were run. The conditions of their practice are indicated in Table II. Groups I and III serve as "rest" controls for the determination of retroactive effects; Group II serves as a control to determine the effects of rest and the immediate effects of task reversal; and Group 5 serves as a control for proactive effects.

All *S*s were male volunteers from a course in elementary psychology who were given two "course points" for each period in the laboratory.

Results

Representative data on standard and reversed tasks.—Group 20 has been chosen to provide a survey of the kinds of data obtained and of the various comparisons to be made. This group ($N = 10$) was given three days of OL, five days of IL, and two days of RL. The principal results are presented in Fig. 1, where mean number of matches and mean number of errors are plotted against trials.

The mean number of matches on OL trial 1 was 12.7. Improvement in performance was fairly rapid and regular. By the last three trials of the third day, the mean was around 35. The initial mean total errors was 10.9, with an increase to 14.8 on trial 2. The total number of errors and the errors on each control remained fairly constant throughout OL.

The mean number of matches on the first trial of IL was 15.5, while the mean total number of errors was 21.2. As compared with the first OL trial, this represents an increase in number of matches and in number of errors, indicating that proactive facilitation as well as proactive interference was operating. As will be seen, the mean number of matches increased steadily during IL, while the mean number of errors decreased, especially during trials 1–30. Of incidental interest, as possible signs of interference effects, were the decrements in number of matches on IL trials of 21, 31, and 41.

The functioning of both proactive facilitation and proactive interference during the first 20 trials of IL is confirmed by Fig. 2. In the left of this figure, the means of matches and the means of total errors for Group 20 on the first 20 trials of IL are compared directly with the means for Group 5 on 20 trials on the reversed task. The means of *matches and errors* for Group 20 are consistently and significantly higher than for Group 5.

Because of the possible relationship, in performance on the Mashburn apparatus, between number of errors and number of matches, each mean of errors was divided by the corresponding mean of number of matches, in order to yield 20 error-

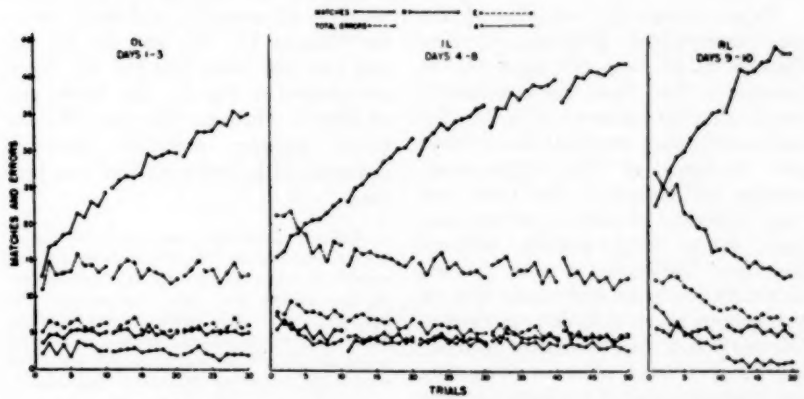

Fig. 1. Matches and error curves during original learning (OL) of *standard* task, interpolated learning (IL) of *reversed* task, and relearning (RL) of *standard* task (Group 20)

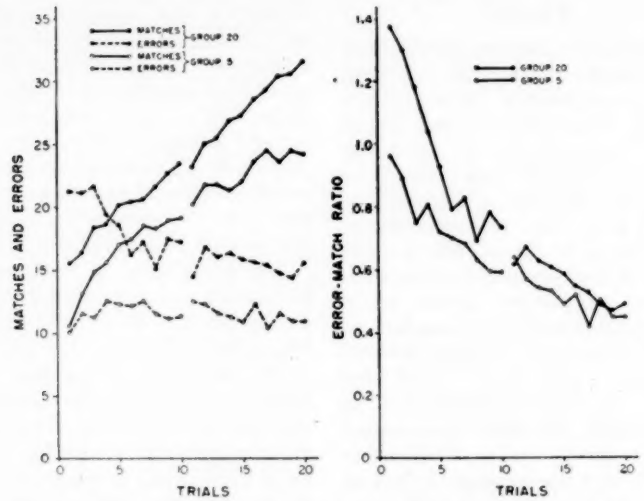

FIG. 2. Matches, errors, and error-match ratios during learning of the *reversed* Mashburn task with and without prior learning of the *standard* task

match ratios for each group of *Ss*. These ratios have been plotted at the right of Fig. 2. With only two exceptions, the ratios for Group 20 were higher than the corresponding ratios for Group 5, and the differences are particularly large during the first few trials.

Representative data on effects of rest and interpolated activities.—Control Groups I, II, and III were run to provide a firm basis for conclusions concerning the amounts of facilitation and interference attributable to different amounts of OL. Exploratory studies had indicated that there was very little loss of skill in performance over fairly long periods without practice. However, it was deemed advisable to obtain systematic data on two groups given different numbers of OL trials and then kept from practicing for five days prior to relearning the original task. Five days rest was chosen because the greatest amount of IL in the experiment was five days.

Part of the data for Group 20 may be compared directly with corresponding data for Groups II and III. Group III had three days of OL and two days of RL separated by five days without practice. Group II was given practice on the *standard* task on 12 consecutive days. The mean number of matches and total errors for Groups II, III, and 20, on the last two OL days and the RL days, are plotted in Fig. 3. Up to the end of Day 3 (through OL trial 30), the three groups received identical amounts of practice on the standard task.

Analysis of variance techniques (6) applied to the matches and error scores on the last several OL trials indicate that the performances of the groups are quite comparable. For example, the *F*-ratio based on the between-groups and within-groups variances of matches scores on OL trial 30 is 1.86, with 2 and 25 *df*, a value well below the .05 level of significance.

Five days intervened between the last trial on Day 3 and the first trial

on Day 9—five days without practice for Group III, five days of additional OL for Group II, and five days of IL for Group 20. The marked differences in performance on Day 9 may be seen in Fig. 3. The large decrement in matches and the large increment in errors for Group 20 are conspicuous, as is the general superiority of the performance of Group II. Of equal importance is the negligible change in the performance of Group III from the end of Day 3 to the beginning of Day 9.

The difference between the means of matches on OL trial 30 and on RL trial 1 for Group III lacks statistical significance ($t=0.75$ with 8 *df*). In contrast, the differences between these means for Group II and for Group 20 are highly significant (t's of 8.81 and 7.95 with 8 and 9 *df*, respectively). The sharp decrease in matches for Group 20 indicates retroactive interference.

Retroactive interference was also shown in Group 20 by the large and significant increase in number of errors ($t=7.74$ with 9 *df*). The change in errors is nonsignificant in Group III ($t=1.19$ with 8 *df*) and in Group II ($t=1.44$ with 8 *df*).

The trends of the matches and error curves for Group 20 on the first day of RL are worthy of special note. The mean number of matches increased rapidly while the mean

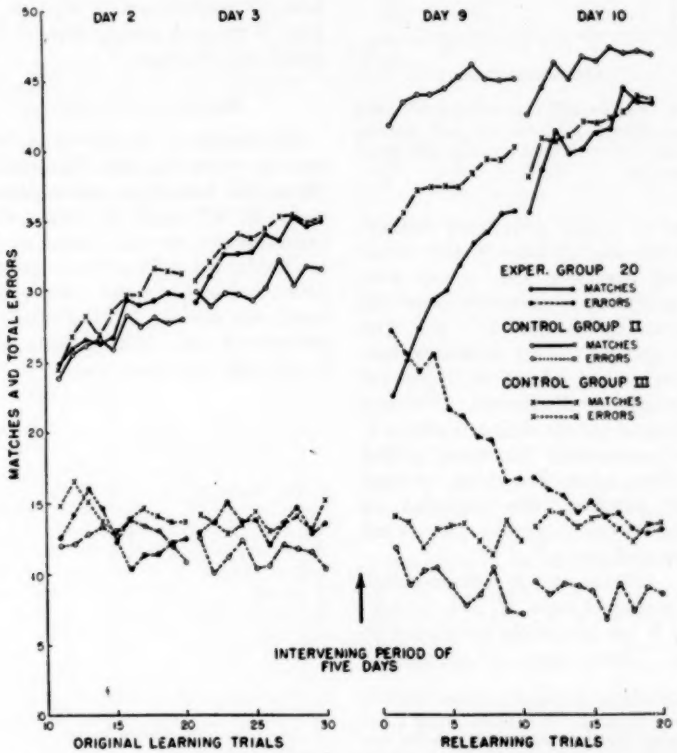

FIG. 3. Matches and total errors during relearning when five days since learning have been filled with rest (Group III), *reversed* task learning (Group 20), or *standard* task learning (Group II)

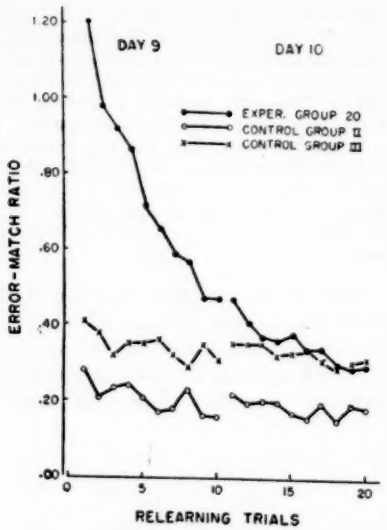

FIG. 4. Error-match ratios during relearning after rest (Group III), *reversed* task learning (Group 20), and continuous standard task learning (Group II)

number of errors decreased rapidly, but at the end of Day 9, the means for both matches and errors were still significantly different from the corresponding means for the two control groups. This indicates that the interference effects had not yet been completely dissipated. The fact that Control III (as well as Control I) retained essentially the same ability to perform after five days without practice justified the omission of separate control groups for each experimental group.[3]

The differences in error-match ratios among Groups II, III, and 20 on Day 9 are strikingly portrayed in Fig. 4. There were no consistent differences between the groups during OL, the error-match ratios on the last OL trial ranging from .33 (Group II) to .43 (Group III). A large error score and a small matches score yields a high ratio which is indicative of inferior performance. It was not until after the second trial of Day 10 that the ratios for Group 20 became essentially the same as those for Group III. The relatively great skill displayed by Group II on Days 9 and 10 is shown by the consistently low ratios. The fact that Group II was markedly superior in performance on Days 9 and 10 provides little or no grounds for attributing any part of the lowered proficiency of Group 20 on Day 9 to such intangibles as "boredom" and "fatigue."

Retroactive Interference

The results of the investigation will now be presented and discussed as if there had been four sub-experiments —A, B, C, and D—with the 12 experimental groups divided as shown in Table III. The three groups in each sub-experiment received the same amount of IL but differed in amount of OL. The amount of RL in all cases was two days.

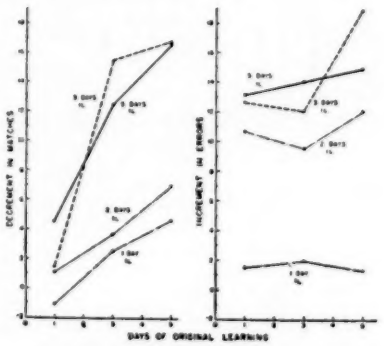

FIG. 5. Decrements in matches and increments in total errors on the first relearning trial as functions of amounts of OL and IL

[3] Results similar in import to those shown in Fig. 3 were obtained for Control Group I and Experimental Group 3 where OL was for five days and IL (Group 3) was for five days. The results have been presented and discussed elsewhere (5).

TABLE III

DECREMENTS IN MATCHES AND INCREMENTS IN ERRORS FROM THE LAST TRIAL OF OL TO THE FIRST TRIAL OF RL AS A FUNCTION OF DIFFERENT AMOUNTS OF OL AND IL

Sub-exp.	Group	*N*	Days of Practice			Mean Decrement in Matches	Mean Increment in Errors	Relative** Decrement in Matches	Relative** Increment in Errors
			OL	IL	RL				
A	15	8	1	1	2	− 1.1*	1.5	−.053*	.109
	18	10	3	1	2	2.5	1.9	.074	.146
	21	10	5	1	2	4.6	1.3	.108	.092
B	30	10	1	2	2	1.1	10.7	.052	.930
	29	10	3	2	2	3.7	9.5	.131	1.173
	31	10	5	2	2	6.8	12.0	.202	1.524
C	16	10	1	3	2	1.4	12.7	.064	.830
	19	8	3	3	2	15.4	12.1	.429	.829
	22	9	5	3	2	16.7	18.8	.410	1.861
D	17	10	1	5	2	4.5	13.2	.186	.898
	20	10	3	5	2	12.4	14.0	.354	1.069
	3	11	5	5	2	16.5	14.8	.429	1.410

* The minus sign indicates an increment instead of a decrement.
** The *relative* decrements and increments were obtained by dividing the *absolute* decrements and increments by the corresponding means on the last trial of OL.

Absolute decrements and increments.—As already indicated, retroactive interference was revealed in Group 20 by decrements in number of matches and by increments in number of errors from the last trial of OL to the first trial of RL. The means of such decrements and increments are listed for the 12 experimental groups in Table III, and they are portrayed graphically in Fig. 5.

As a first step in evaluating the significance of the absolute decrements and increments given in Table III, it was necessary to investigate the initial comparability of the three groups within each sub-experiment. Simple analysis of variance techniques were used for this purpose and were applied to the matches scores and error scores obtained by the three groups in each sub-experiment on trial 10 of OL, since all groups received identical treatment through that trial. All eight *F*-ratios—two for each sub-experiment—fell below the .05 level of significance, and the null hypothesis cannot be rejected. Real differences appearing later in either IL or RL may be attributed to differences in the treatment given subsequent to OL trial 10.

Analysis of variance was used in relation to the means of absolute decrements in matches shown in Table III. The procedure was to utilize the means along with the decrements for individual *S*s in the three groups composing each sub-experiment, and obtain an *F*-value by dividing the between-groups variance by the within-groups variance. The four *F*-values secured in this way were all significant beyond the .01 level.[4]

In addition to being significantly different within each of the sub-experiments, the decrements in number of matches from the last trial of OL to the first trial of RL increased in magnitude as the number of days of OL increased. In other words, with the amount of IL held constant at either 1, 2, 3, or 5 days, the decrements in matches from the last trial of OL to the first trial of RL

[4] For Experiment A, $F=6.34$ with 2 and 25 *df*; for B, $F=6.69$ with 2 and 27 *df*; for C, $F=19.15$ with 2 and 24 *df*; and for D, $F=13.28$ with 2 and 28 *df*.

always *increased* with the amount of OL.

The negative "decrement" in matches of −1.1 shown in Table III for Group 15 represents an increment instead of a decrement. This increment suggests that a small amount of OL followed by a small amount of practice on a similar interpolated task may yield more facilitation than interference. There is, perhaps, a minimum level of learning which must be exceeded before one motor task will interfere with another.

The *t*-test for unrelated measures (6) was employed to evaluate the differences between the means of decrements in matches within each sub-experiment. The obtained values of *t* and their levels of significance are given in Table IV. It will be noted that, regardless of the amount of IL (that is, regardless of the sub-experiment involved), the decrement in matches is greater for five days of OL than for one day of OL. The differences in all cases are significant beyond the .01 level. With one day of IL (Sub-experiment A), the difference between the decrements for one and three days of OL is significant between the .02 and .05 levels, while with two days of IL (Sub-experiment B), the difference for one and three days of OL falls below the .10 level. In contrast, with three days and also with five days of IL, the differences between the matches decrements arising from one and three days of OL are significant beyond the .01 level. Regardless of the amount of IL, there are no significant differences between three and five days of OL.

The lack of statistical significance of some of the differences between matches decrements within the sub-experiments should not be allowed to obscure the probable importance of the trends shown by the four curves in Fig. 5. In every sub-experiment (that is, for each of the four amounts of IL), there was a progressive increase in the size of the absolute decrement in matches as the amount of OL was increased.

The absolute increments in number of errors from the last OL trial to the first RL trial tended to be greater for the larger amounts of IL, but they failed to show systematic tendencies either to increase or decrease with changes in the amount of OL (see

TABLE IV

SIGNIFICANCE OF DIFFERENCES BETWEEN THE MEANS OF DECREMENTS IN MATCHES FROM THE LAST TRIAL OF OL TO THE FIRST TRIAL OF RL

Sub-exp.	Groups Compared	Amounts of OL in Days	*t*	*df*	Level of Significance
A	15 & 18	1 & 3	2.22	25	$.05 > P > .02$
	15 & 21	1 & 5	3.52	25	$.01 > P > .001$
	18 & 21	3 & 5	1.38	25	$.20 > P > .15$
B	30 & 29	1 & 3	1.67	27	$.15 > P > .10$
	30 & 31	1 & 5	3.65	27	$.01 > P > .001$
	29 & 31	3 & 5	1.99	27	$.10 > P > .05$
C	16 & 19	1 & 3	4.95	24	$.001 > P$
	16 & 22	1 & 5	5.58	24	$.001 > P$
	19 & 22	3 & 5	0.45	24	$.70 > P > .60$
D	17 & 20	1 & 3	3.26	28	$.01 > P > .001$
	17 & 3	1 & 5	5.09	28	$.001 > P$
	20 & 3	3 & 5	1.74	28	$.10 > P > .05$

Table III). Analysis of variance techniques (6) were applied to the error scores for the three groups of *Ss* within each sub-experiment. The *F*-values all fell below the .05 level of confidence. It must be assumed, therefore, that amount of retroactive interference, as represented by absolute increments in errors, is independent of the amount of OL.[5]

Relative decrements and increments.—Even though the absolute increments in errors appeared to be unrelated to the amount of OL, the relative increments (shown in the last column of Table III) are, with one important exception, greater for the larger amounts of OL. The exception is Group 21. The relative increments show that there is probably some tendency for an increased amount of OL to induce a relatively larger number of errors at the outset of RL.

The relative decrements in matches are generally in agreement with and lead to the same conclusions as do the absolute decrements. There is a single discrepancy—the inversion for Group 19. The trend of the relative values lends further support for the view that retroactive interference is importantly related to the amount of OL.

Proactive Facilitation and Proactive Interference

As used in this paper, the term proactive facilitation refers to improvement in performance (increase in matches, decrease in errors) during IL, which may be attributed to OL. The term proactive interference refers to deterioration of performance during IL, which is attributable to OL. The improvement or deterioration must be determined with respect to performance data obtained from *Ss* who began practice on the interpolated (reversed) task.

[5] Despite their seeming independence of amount of OL, absolute increments in errors from the last OL trial to the first RL trial were found to be importantly related to the amount of IL. This point will be covered in (8).

TABLE V

DIFFERENCE BETWEEN FIRST-TRIAL PERFORMANCE ON THE REVERSED MASHBURN TASK AS A FUNCTION OF DEGREE OF OL OF THE STANDARD TASK

OL Condition			Mean Differences on First IL Trial	
Sub-exp.	Group	OL Days	Matches*	Errors*
A	15	1	3.3	10.9
	18	3	4.6	12.8
	21	5	6.6	14.9
B	30	1	4.6	10.0
	29	3	4.2	10.7
	31	5	6.7	13.5
C	16	1	5.2	10.0
	19	3	5.5	8.9
	22	5	4.6	13.9
D	17	1	6.8	11.4
	20	3	4.9	11.1
	3	5	5.4	9.3

* The differences, all positive, are in favor of the experimental groups.

The simultaneous functioning of facilitative and interfering factors during IL on the Mashburn apparatus can be demonstrated quite conclusively with the results from the present experiment. A comparison has already been made in Fig. 2 between the data for the first 20 trials of IL for Group 20 and the data for 20 initial reversed trials given Group 5. Group 20 made significantly greater numbers of matches and significantly greater numbers of errors indicating, respectively, proactive facilitation and proactive interference.

Comparisons between mean matches and errors on the first trial for Group 5 and on the first trial for all experimental groups are given in Table V.

The *t*-test for unrelated measures (6) was employed to evaluate each of these differences. All but two of the differences between the matches means were found to be significant at better than the .01 level of confidence. The exceptions were the differences for Groups 15 and 18, which were found to lie, respectively, around the .05 and .02 levels. It is reasonable to assume, therefore, that the performance of all of the experimental groups at the outset of IL was affected by some amount of proactive facilitation.[6]

As already seen in Fig. 2, Group 20 showed superiority in IL performance over Group 5, not only on trial 1 but also on succeeding trials. The other experimental groups were also found to show superiority over Group 5 during succeeding IL trials. This means that the facilitative effects were not fleeting in character but had a degree of permanence.

The unmistakable functioning of facilitation in the IL performance of the experimental groups led to the conjecture that increasing amounts of OL would produce increasing amounts of facilitation. There is evidence both for and against this notion. For one analysis, the 12 sub-experimental groups were placed in three larger groups depending on whether they received 1, 3, or 5 days of OL. The means of number of matches and of total number of errors for the three combined groups on each of the first ten trials of IL were computed. A plot of these means appears in Fig. 6, where solid lines connect means of matches and dashed lines connect means of errors. The curves for the different groups may be identified by the numerals adjacent to them.

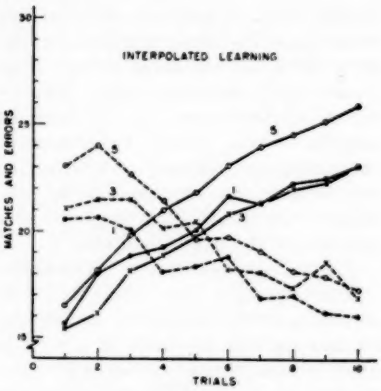

FIG. 6. Matches and total errors during the first ten trials of IL after 1, 3, and 5 days of OL

The matches curves for *S*s having one and three days of OL show common trends (except on the second trial), but the matches curve for *S*s having five days of OL is consistently above the other two. The *F*-test (6) was used to evaluate the hypothesis that the three groups of matches scores on the first IL trial came from a single normal population of scores. The hypothesis is quite tenable ($F = 1.11$ with 2 and 113 *df*). A similar test involving scores on IL trial 10 yielded an *F*-value of 5.79, which, with 2 and 113 *df*, is significant beyond the .01 level of confidence. The differences between the mean on trial 10 for *S*s having five days of OL and the means for *S*s having one and three days of OL are significant beyond the .01 level, as found by applying the *t*-test for unrelated measures (6).

The data therefore lend support to the view that the amount of proactive facilitation is not independent of the

[6] These comparisons assume that the *S*s of Group 5 and of the experimental groups were randomly chosen from the population and were, therefore, homogeneous in the ability to perform. This assumption seemed warranted by the way the *S*s for each group were obtained and assigned to groups.

number of OL trials. However, the nature of the relationship is obscure. There is no obvious explanation for the lack of difference in the performance of *Ss* having one day of OL and those having three days of OL. Perhaps an interaction of facilitation and interference was responsible. It is apparent from Fig. 6 that the matches means for *Ss* having five days of OL became conspicuously greater than the means for the other *Ss*, after the means of errors (and, inferentially, the magnitude of the interfering effects) had decreased.

Another analysis bearing on proactive facilitation utilized the data for the three groups in each of Subexperiments B, C, and D. The purpose was to determine whether different amounts of OL led to differences in performance beyond the tenth IL trial. The analysis for Sub-experiment C will be discussed in some detail. Matches and error curves for Groups 16, 19, and 22 on certain of the OL, IL, and RL trials are presented in Fig. 7. In the left graph, curves are shown for the first ten and last ten trials of OL. The two curves labeled *5* (solid lines for matches, dashed lines for errors) are for Group 22, the group receiving five days of OL. All other curves in Fig. 7 like these two, are for the same group. The curves labeled *3* are for Group 19, and are based on data for the third day of OL. The curves labeled *1* are for Group 16, and summarize data for the first and only day of OL. To facilitate comparisons, the OL curves for Group 16 are shown twice. The group had only ten OL trials, so these constituted the first ten as well as the last ten trials for the group.

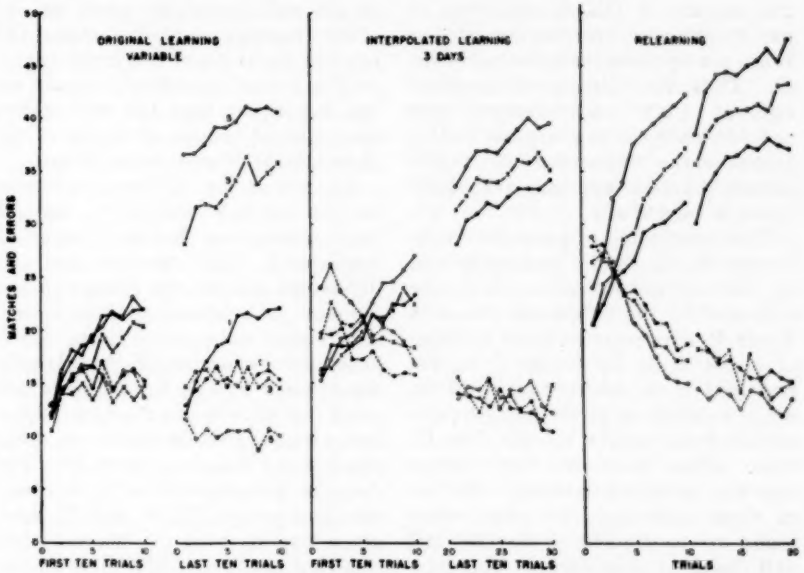

FIG. 7. Matches and total errors during first and last ten trials of OL, first and last ten trials of IL, and RL, after 1, 3, and 5 days of OL

The middle graph of Fig. 7 summarizes the performance of the three groups in Sub-experiment C on the first ten and last ten trials of IL. The groups all received three days of IL. The three matches curves for the last ten trials of IL fall conspicuously in order from the greatest to the least amount of OL. The three sets of means and the individual scores on which they are based were subjected to a trend analysis of the type originally proposed by Lindquist (7, Case 11, p. 76) and later extended by Kogan (3). The value of *F* bearing on the grand means of the three groups is not large enough to permit a confident rejection of the hypothesis that the true differences among them is zero, but it approaches the .05 level.

The conclusion suggested by the foregoing analysis, that the amount of proactive facilitation depends upon the amount of OL_2, is supported in part by the results for Sub-experiment B but not by those for Sub-experiment D. Thus, the evidence is inconsistent and at times contradictory, and provides no basis for a firm conclusion. However, the conjectured relationship seemed plausible and may eventually prove to be correct.

The functioning of proactive interference during the IL performance of the 12 experimental groups is clearly indicated by the differences shown in Table V between the mean number of errors made by Group 5 on the first trial on the reversed task and the mean numbers of errors made by the experimental groups on the first IL trial. When evaluated with the *t*-test for unrelated measures (6), all of these differences in errors were found to be significant beyond the .001 level of confidence. The persistence of the interfering effects was shown in connection with the data of Fig. 2, where the performances of Groups 5 and 20 were compared over 20 trials. The interfering effects were also found to persist in the performances of the other experimental groups.

It seemed reasonable to suppose that proactive interference, as revealed by increases in number of errors, would be positively related to the amount of OL. The evidence on proactive interference is suggestive but inconclusive. As shown in Fig. 6, the means of errors on the first four trials of IL are greatest for *Ss* having five days of OL and least for those having one day. Except on trial 6, the *Ss* having one day of OL made the fewest errors, on the average, on all IL trials. Simple analysis of variance was applied to the three sets of error scores for the first IL trial. The resulting *F*-value of 2.33 (with 2 and 113 *df*) falls below the .05 level, so the null hypothesis could not be safely rejected. Nevertheless, the relative positions of the error curves in Fig. 6 cast considerable doubt on the hypothesis that the true differences among the grand means of the three groups of error scores is zero.

As seen in Fig. 7, the error curves for the last ten trials of IL for the three groups in Sub-experiment C overlapped. This indicates that any differences among the groups in the amount of proactive interference which may have arisen from differences in amount of OL had largely disappeared by the beginning of the third day of reversed practice. Confirmation of this finding came from the data for Sub-experiment D. The data for Sub-experiment B, in which the three groups (29, 30, and 31) were given two days of IL, showed that small but consistent differences among the means of errors persisted almost to the end of the second day. Group

31, which had five days of OL, made more errors on the second day of IL than did Group 30, except on the last trial. The latter group received only one day of OL.

The curves in the right-hand graph of Fig. 7 reveal the general course of RL on the Mashburn apparatus when different amounts of OL were provided prior to three days of IL. They have no bearing on proactive effects but enable the reader to make direct comparisons of the changes in means of matches and means of errors over the two days of RL; and these changes may be compared with those occurring during the other two phases of learning.

Summary

The modified Mashburn apparatus was utilized in a systematic study of the role of proactive and retroactive facilitation and interference in the performance of two closely similar but different complex tasks.

Twelve groups of *Ss* first acquired skill on the *standard* Mashburn task for 10, 30, and 50 trials, then learned the reversed task for 10, 20, 30, or 50 trials, and finally relearned the original task for 20 trials. Two control groups, after 30 and 50 trials of OL, rested for five days, and then relearned. A third control group practiced continuously on the standard task for 120 trials. A fourth control group began on the reversed task and practiced for 20 trials. All trials were 2 min., with 15-sec. intertrial rests, ten trials per day.

Facilitation during interpolated learning and relearning was measured by increases in number of matches and/or decreases in number of errors, while interference was revealed by decreases in matches and/or increases in errors. The following conclusions were reached:

1. The acquisition of skill on the standard Mashburn task facilitated and also interfered with the subsequent learning of the reversed task. The signs of both facilitation and interference were clearly present at the outset of reversed practice and persisted in recognizable form through several trials.

2. Contrary to expectations, the amounts of proactive facilitation and interference displayed during IL were not found to be positively related to the amount of OL on the standard task. However, the evidence was equivocal, indicating that such a relationship may actually hold.

3. Only one of the 12 experimental groups yielded clear evidence of the functioning of retroactive facilitation. The *Ss* in this group were given only one day of OL and one day of IL. It is believed that retroactive facilitation was probably operating for all experimental groups during RL but was counteracted and obscured by the relatively large amounts of retroactive interference.

4. The experimental groups, especially those receiving three or five days of OL and IL, showed conspicuous signs of retroactive interference during the early stages of RL. In most cases, there were highly significant decrements in number of matches and/or highly significant increments in number of errors from the last trial of OL to the first trial of RL. The absolute decrements as well as the relative decrements in number of matches were positively related to the amount of OL. The absolute increments in number of errors were seemingly unrelated to the amount of OL. All amounts of OL, regardless of the amount of IL, led to absolute increments in errors, but the increments tended to be of the same general magnitude for any specified amount of IL. The relative increments in errors were dependent to some extent upon the amount of OL.

5. Results for the control groups showed that the decreases in matches and increases in errors displayed by the experimental groups at the outset of RL were the result of retroactive interference and not the result of the mere passage of time from the last OL trial to the first RL trial, nor of such intangible factors as boredom and fatigue.

(Manuscript received March 27, 1950)

REFERENCES

1. BUXTON, C. E., AND HENRY, C. E. Retroaction and gains in motor learning: I. Similarity of interpolated task as a factor in gains. *J. exp. Psychol.*, 1939, **25**, 1–17.
2. BUXTON, C. E., AND GRANT, D. A. Retroaction and gains in motor learning: II. Sex differences, and a further analysis of gains. *J. exp. Psychol.*, 1939, **25**, 198–208.
3. KOGAN, L. S. Analysis of variance: Repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.
4 LEWIS, D. Positive and negative transfer in motor learning. *Amer. Psychologist*, 1947, **2**, 423. (Abstract.)
5. LEWIS, D., AND SHEPHARD, A. H. Devices for studying associative interference in psychomotor performance: I. The modified Mashburn apparatus. *J. Psychol.*, 1950, **29**, 35–46.
6. LINDQUIST, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.
7. LINDQUIST, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.
8. McALLISTER, DOROTHY E., AND LEWIS, D. Facilitation and interference in performance on the modified Mashburn apparatus: II. The effects of varying the amount of interpolated learning. *J. exp. Psychol.*, 1951, **41**. (In press.)
9. McGEOCH, J. A. The influence of degrees of learning upon retroactive inhibition. *Amer. J. Psychol.*, 1929, **41**, 252–262.
10. MASHBURN, N. C. Mashburn automatic serial action apparatus for detecting flying aptitude. *J. Aviat. Med.*, 1934, **5**, 155–160.
11. MELTON, A. W. (Ed.) *Apparatus tests.* Army Air Forces Aviation Psychology Program Research Report No. 4. Washington, D. C.: U. S. Govt. Printing Office, 1947.

TEMPORAL GRADIENTS OF RESPONSE STRENGTH WITH TWO LEVELS OF MOTIVATION[1]

BY GERALD ROSENBAUM[2]

State University of Iowa

Several investigators have observed that the repeated presentation of stimuli at regular time intervals may produce correspondingly spaced fluctuations in the strength of a response, its strength increasing with proximity to the customary time of stimulation. This phenomenon was originally reported by Pavlov (8, p. 41) in his studies of the salivation reflex. In view of his finding that the reflex occurred periodically at the regularly conditioned time interval in the absence of any other conditioned stimulus, the procedure has been characterized as temporal conditioning. The progressive increase in response strength resulting from the presentation of stimuli at successively shorter temporal distances from the usual time of stimulation is described as a temporal gradient.

In studies concerned with the hunger cycle of activity in rats, Richter (10) and Skinner (11, p. 366) have observed a gradient of accelerated spontaneous activity during the interval prior to the usual feeding time. Brown (1) has shown that rats previously shocked at regularly recurring 12-sec. intervals, exhibited a gradient in the amplitude of their unconditioned responses to shock, when they received shocks at successively shorter intervals. Temporal conditioning has also been demonstrated with human *Ss*. A study conducted in Hull's laboratory (3) revealed that periodic stimulation by an electric shock resulted in three successive galvanic reactions after the stimulus sequence had been discontinued. Mowrer (7) has subsequently obtained a gradient of response latencies with human *Ss* by having them release a key in response to a tone presented at varied temporal distances from the previous training interval.

The present experiment was designed to investigate a temporal gradient in an instrumental conditioning situation and to determine the effect of reduced motivation upon the slope of the gradient. The general procedure employed was as follows: rats were trained to depress a bar every 60 sec. for a food reward, the latency of the bar-pressing response being recorded. The rats were then presented with the bar at test intervals shorter and longer than 60 sec. in order to determine whether the variations in response latency would coincide with temporal distance from the training interval. Two groups of animals under two different drive conditions were used to test for the effects of drive strength.

Apparatus

The modified Skinner Box described by Perin (9) was employed, except that the 12-in. compartment was used without any restraining inner compartment. In addition, a 2-v. radio dial lamp (approximately 1 c.p.) was situated above the food tray. The light from this lamp

[1] This experiment was submitted to the faculty of the State University of Iowa in partial fulfillment of the requirements for the degree of Master of Arts. The writer is deeply indebted to Dr. J. S. Brown for his advice and encouragement. An abstract of this paper was presented at the 1948 meeting of the Midwestern Psychological Association.

[2] Now at Wayne University.

served to illuminate the box and may have helped to orient *S* toward the bar. The bar could be inserted and withdrawn as desired through a slit in the rear panel of the box. The response measure was the time taken by *S* to depress the bar following its insertion by *E*, using a lever on the outside of the box. A microswitch, located within the plane of rotation of the lever arm, activated a Standard Electric Timer in a locking circuit when the bar was presented. A downward pressure on the bar of 17 gm. was sufficient to break the circuit, stopping the clock. Simultaneously, the circuit to the food-release mechanism was made, releasing a pellet of food to the cup beneath the horizontal bar. The pellets were cylindrical in shape and weighed about 50 mg. A stop watch was used to measure the intervals between bar presentations.

Procedure

Preliminary training.—Sixteen male hooded rats, four to five months old, were fed 8 gm. of Purina Dog Chow biscuits in individual feeding cages at the same hour each day for a week, in order to establish a definite feeding rhythm. They were later run at their customary feeding time. Water was present in the feeding and living cages at all times.

Habituation to the food-release mechanism (Day 1).—Each *S* was placed in the experimental box 22 hr. after its last feeding. Two pellets were already in the food tray. While the second pellet was being eaten, *E* closed a contact which caused another pellet to drop into the food tray. After an interval of 60 sec., another pellet was delivered. This procedure was continued at 60-sec. intervals until 48 pellets had been consumed. In the course of this training, *S* learned to respond to the click of the food-release mechanism by going immediately to the food tray and eating the pellet. It also became accustomed to the 60-sec. interval between successive eatings. Following this training, *S* was placed in a feeding cage for 1 hr., where it was permitted to consume the remaining portion (5.6 gm.) of its daily ration of 8 gm.

Acquisition of the bar-pressing response (Day 2).—A small quantity of food mash was smeared on the panel above the bar and on the bar itself prior to *S*'s entry into the box. The *S*s were again motivated by a 22-hr. hunger drive, and the bar was withdrawn at the time *S* was placed in the compartment. After an interval of 60 sec., the bar was inserted into the compartment. In attempting to obtain the food which had been smeared on the bar, *S* depressed the bar sufficiently to activate the food-release mechanism, dropping a pellet into the food tray. The bar was withdrawn immediately, and *S* ate the pellet. Sixty sec. following the first introduction of the bar, it was again presented and the procedure repeated. When *S* had completed 48 bar-depressions, it was removed from the box and placed in a feeding cage.

Further conditioning (Day 3).—The procedure employed on the second day was repeated, except that no food mash was present in the compartment. Each *S* was again operating under the 22-hr. training drive and received 48 reinforcements at the standard 60-sec. interval.

Testing series I (Day 4).—The *S*s were divided into Groups A and B with eight *S*s, equated for litter, in each group. For Group A, the 22-hr. training drive was maintained, but the drive level of Group B was reduced by feeding them 4 gm. of Purina Dog Chow just prior to being placed in the experimental box. Following five reinforcements at the standard interval, a series of test trials was given in which the bar was introduced at the following test intervals: 15, 30, 45, 60, 75, 90, 105, and 120 sec., though not necessarily in that order. Between each pair of test trials, five presentations of the bar at the standard interval occurred so that disruption of the established temporal rhythm by the interpolation of irregular test intervals was minimized. Each *S* thus received 48 reinforcements under the testing conditions. The order of presentation of the test intervals was varied systematically to preclude serial position effects.

Renewed conditioning (Day 5).—Groups A and B received 48 reinforcements at the standard interval motivated by the 22-hr. training drive.

Testing series II (Day 6).—The testing procedure was the same as on Day 4, but the drive state was reversed for the two groups. The same order of presentation was maintained for every animal. Thus each animal could be compared with itself for the two drive conditions, and practice effects were equated.

Results

The strength of the response tendency at each of the test intervals is shown in Fig. 1. The data for all 16 *S*s under both drive conditions were combined in computing the ordinate values. By taking the average latency of the two test trials at each interval, the most stable estimate of the response tendency was obtained. Another way of treating the data, previously used by Mowrer (7), is to

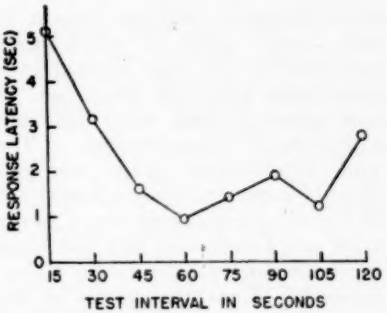

FIG. 1. Composite curve showing mean latency of response as a function of the test interval

express the latency at each test interval as a percentage of the mean of the three previous trials at the standard interval. The general shape of the curve when the data are plotted in this fashion closely approximates that for the absolute latencies.

From Fig. 1 it is apparent that the strength of the tendency to respond, expressed in terms of the length of time it takes the rat to depress the bar, is greatest at the standard 60-sec. interval. When the bar is presented at intervals shorter than the one to which the *Ss* have been trained, they take progressively longer to respond with increasing temporal distance from the training interval. The *t*'s for related measures computed between the means for each of these short test intervals and the mean at the standard interval are all significant at better than the .01 level of confidence; the values of *t* for the differences between points *15* and *30* and points *30* and *45* are significant at better than the .05 level of confidence.[3]

[3] A control group of four *Ss* was subjected to the same experimental procedure, but it was always tested at the 60-sec. interval. The mean response latencies obtained for this group on the respective test trials were always significantly faster than the means for the experimental *Ss* at the short test intervals.

When the standard interval is passed and the bar inserted at increasingly greater temporal intervals, the response latency tends to increase. However, the differences between any of these longer intervals and the 60-sec. interval are not statistically significant since these latencies were extremely variable. Thus it appears that there is a gradient of response latencies which decreases in height as the conditioned temporal stimulus is approached in time. If the habitual point of stimulation is passed and the response is not permitted to occur, the response latencies tend to remain at the same level as at the conditioned interval, perhaps increasing slightly with temporal distance from the training interval.

In Fig. 2 the effect of the two drive states upon the gradient of response strength is shown. The curves presented suggest that the reduced drive condition tends to steepen the slope of the gradient, but a statistical analysis of the difference between the two slopes reveals that it is not significant. However, when Groups A and B are compared for the trials on the first testing day (Day 4) only,

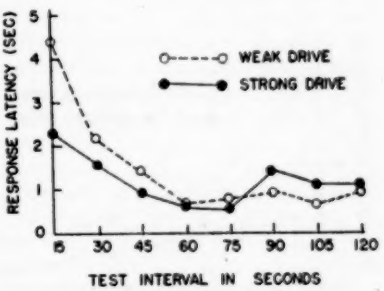

FIG. 2. Comparative median latencies for the two drive conditions. The strong drive condition consists of the responses obtained under the 22-hr. hunger drive, while the weak drive condition consists of responses obtained following the ingestion of 4 gm. of food.

the difference at the 15-sec. interval is significant at better than the .05 level of confidence. Similarly, a comparison of the mean latencies on all interspersed training trials on Day 4 indicates that the reduced drive condition tends to increase the response time. The effect of additional habituation to the experimental situation by the second testing day (Day 6) appears to have obliterated the differences between the two drive conditions for the combined data. Furthermore, analysis of the individual records discloses that certain animals in Group B gave consistently longer latencies even under the strong drive state, which would also tend to obscure any difference. A comparison between Day 4 and Day 6, which would indicate the effects of differential practice, is not presented as there was no clear-cut difference between the two curves, although the latencies for the standard interval on Day 4 tend to be consistently longer than those on Day 6.

In view of the suggestive but inconclusive results obtained from the two drive conditions, a subsidiary experiment was conducted in which a greater reduction of the drive state was attempted. Sixteen different *Ss* were used, eight of which were motivated by the 22-hr. drive and the other eight were now fed 10 gm. of food prior to their entry into the experimental compartment. The same experimental design was employed but the procedure was modified so that the animals were tested on the third and fourth days and test trials were run only on the 15- to 60-sec. intervals. From the results of the previous experiment, it was predicted that a more reliable picture of the effect of drive reduction could be obtained by increasing the difference between the drive states of the two groups and decreasing the amount of practice.

The curves for the two groups are presented in Fig. 3 and support the prediction derived from the previous experiment. A *t*-test applied to the distribution of differences between the latencies at the 15-sec. interval and those at the 60-sec. interval, reveals that the mean difference for the weak drive condition is significantly greater than the mean difference for the strong drive condition at better than the .01 level of confidence. This finding indicates that reduced motivation tends to steepen the slope of the temporal gradient of response strength.

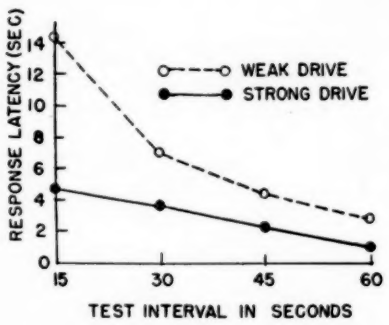

FIG. 3. Comparison of mean latencies for a greater difference between the two drive conditions (Experiment II).

DISCUSSION

One possible explanation of the gradient phenomenon obtained in this experiment is Mowrer's concept of expectancy (7). He suggests that anticipation of recurrent events builds up an increasing amount of tension prior to a repeated stimulus event. The intensity of the state of tension is assumed to be quantifiable in terms of such measures of the strength of the response as latency and magnitude. According to this hypothesis, the

gradient of response latencies observed in the present experiment would be considered a direct representation of the course of expectancy. This formulation obviously does not attempt to account for the progressive increase in tension or expectancy associated with greater proximity to the repeated stimulus event.

An alternative explanation from which the temporal gradient can be derived involves the concept of secondary reinforcement. Since the presentation of the bar is followed by a goal response resulting in immediate reinforcement, the bar presentation itself may serve as a secondary reinforcer for the responses of orienting toward the bar, which immediately precede it. In Spence's recent account of delayed reward learning (12), he postulates immediate differential secondary reinforcement of responses associated with internal stimuli and suggests that the secondary reinforcing properties of a stimulus pattern generalizes to the stimulus patterns preceding it in time according to some gradient. In terms of this hypothesis, the strength of the bar-orienting responses within the 60-sec. temporal sequence would be inversely related to their temporal distance from the customary time for the reinforcing bar presentation. The progressively longer latencies at the shorter temporal intervals would then be a function of the decreasing probability of orienting responses occurring at these intervals.

Since orienting responses were not actually measured in this experiment, a more direct explanation of the present findings may be derived from the principle of stimulus generalization without invoking the concept of secondary reinforcement. In addition to the visual and auditory stimuli accompanying each presentation of the bar, the standard temporal interval between successive presentations may become a conditioned stimulus for the bar-pressing response. The actual stimulus cue associated with the temporal interval may be some particular phase of the internal activity sequence, such as a segment of the trace of the preceding response or a stage of the digestive process following the ingestion of the last pellet. Stimulus generalization would then occur along this dimension of internal stimuli correlated with the temporal sequence, and the stimuli most similar to those at the conditioned interval would elicit the strongest responses. For intervals greater than 60 sec. the conditioned stimulus would have already occurred and *S* would be set to respond. Thus, there was no appreciable difference in response latencies for these longer intervals. Extinction could not take place since every test trial was followed by five conditioning trials.

This latter finding is at variance with the results of Brown (1) and Mowrer (7). Brown obtained a sharp drop in response amplitude at the interval immediately following the standard, and a gradual strengthening of the response tendency as the test interval approached a point twice as large as the standard. The difference in results is probably attributable to the fact that Brown was dealing with the facilitating effect of temporal conditioning upon the magnitude of an unconditioned response to shock. The internal cue at the standard interval which intensified the animal's jumping response may also have served as a cue for an incompatible relaxing response when the test shock failed to appear. Mowrer obtained a gradual increase in reaction times for test intervals longer than the standard. However, the slope of the

curve was much flatter than that obtained for the shorter intervals and more closely approximates the present findings, the difference probably arising as the result of verbal cues which served to mediate the speed of the response on the longer intervals for human *Ss*.

By considering the temporal gradient of response latencies as a reflection of stimulus generalization upon the curve of response strength, it is possible to deduce the effects of different drive states upon the slope of the gradient from Hull's theoretical system (**5**). Reaction latency ($_{s}t_{R}$) is a negatively accelerated decreasing function of effective reaction potential ($_{s}\bar{E}_{R}$). If drive (D) is held constant, then the gradient of latencies presented here reflects successive values of effective habit strength ($_{s}\bar{H}_{R}$), which is a function of stimulus generalization. Since D is assumed to multiply all $_{s}\bar{H}_{R}$ values, greater values of $_{s}\bar{E}_{R}$ would be predicted for the strong drive gradient than for the weak drive gradient.[4] However, reaction latency is assumed to be a negatively accelerated function of $_{s}\bar{E}_{R}$. Thus, the greater the values of $_{s}\bar{E}_{R}$, the smaller would be the differences between the predicted latencies for these $_{s}\bar{E}_{R}$ values. This means, in effect, that the latency gradient of generalization becomes flatter under strong drive conditions and steeper under weak drive conditions. The results presented in Fig. 3 illustrate this prediction.

This finding supports the results obtained by Brown (**2**) for a generalization gradient of intensities of illumination and relates to previous evidence by Hull (**4**) and Miller and Miles (**6**) that reducing the drive tends to raise the slope of the gradient of running speed. The conclusion that lowered motivation steepens the slope of the generalization gradient applies only to response latency in Hull's system. Further investigation in this area is required using other measures of response strength.

[4] Increases in D should steepen as well as elevate the generalization gradient of $_{s}\bar{E}_{R}$. Consequently, any change in the slope of the gradient of latencies under strong drive conditions depends upon the amount of drive increase and the actual slope of the negatively accelerated curve relating $_{s}\bar{E}_{R}$ values and reaction latency. For the curve proposed by Hull, increases in drive will result in a flattening of the gradient of latencies in most instances, the exception being cases in which the $_{s}\bar{E}_{R}$ values and the amount of drive increase are both very small.

Summary

1. The experiment tested for the presence of a temporal gradient of response strength in an instrumental conditioning situation, and for the effect of two different drive conditions upon the slope of the gradient.

2. Sixteen rats, matched for litter, were trained to depress a bar in a Skinner Box, and the response latencies were recorded. The bar was inserted periodically every 60 sec. for 96 trials, so that a temporal rhythm was established. The *Ss* were then tested at the 60-sec. interval and at intervals shorter and longer than 60 sec. Half of the *Ss* were tested under the 22-hr. training drive, and the other half were tested under a reduced drive condition. The drive states were reversed for the two groups following 192 trials, and *Ss* were again tested at the varied temporal intervals.

3. In the tests, the shortest latencies occurred at the conditioned 60-sec. interval. For intervals shorter than 60 sec., latency becomes progressively longer with increasing temporal distance from the conditioned interval. For intervals longer

than 60 sec., latency was not significantly increased.

4. No significant differences were obtained between the gradients of response strength for the two drive conditions, although there is some evidence that the reduced drive increased response latency, particularly at the shortest test interval.

5. A second experiment, utilizing a greater difference in drive between the two groups, revealed a significant difference between the slopes of the two gradients for the 15- to 60-sec. intervals.

6. An explanation of the temporal gradient is proposed in terms of the generalization of internal stimulus cues, and the steeper gradient of response latencies for the reduced drive condition is deduced from Hull's theoretical system.

(Manuscript received March 27, 1950)

References

1. Brown, J. S. A note on a temporal gradient of reinforcement. *J. exp. Psychol.*, 1939, **25**, 221–227.
2. Brown, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.
3. Hull, C. L. Learning: II. The factor of the conditioned reflex. In C. Murchison (Ed.), *Handbook of general experimental psychology*. Worcester, Mass.: Clark Univ. Press, 1934.
4. Hull, C. L. The rat's speed-of-locomotion gradient in the approach to food. *J. comp. Psychol.*, 1934, **17**, 393–422.
5. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
6. Miller, N. E., and Miles, W. R. Effect of caffeine on the running speed of hungry, satiated, and frustrated rats. *J. comp. Psychol.*, 1935, **20**, 397–412.
7. Mowrer, O. H. Preparatory set (expectancy)—some methods of measurement. *Psychol. Monogr.*, 1940, **52**, 1–43.
8. Pavlov, I. P. *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. (Trans. and ed. by G. V. Anrep.) London: Oxford Univ. Press, 1927.
9. Perin, C. T. A quantitative investigation of the delay of reinforcement gradient. *J. exp. Psychol.*, 1943, **32**, 37–51.
10. Richter, C. P. Animal behavior and internal drives. *Quart. Rev. Biol.*, 1927, **2**, 307–343.
11. Skinner, B. F. *The behavior of organisms*. New York: D. Appleton-Century, 1938.
12. Spence, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, **54**, 1–8.

THE EFFECTS OF DIFFERENTIAL REWARDS ON DISCRIMINATION REVERSAL LEARNING BY MONKEYS [1]

BY DONALD R. MEYER

The Ohio State University

There are few kinds of empirical evidence that have as tenaciously resisted incorporation by drive-reduction theory as that which demonstrates performances not to be related in any simple way to the physical characteristics of reinforcing agents. Changes in the facilitative properties of rewards that accompany qualitative or quantitative shifts have been cited frequently in support of the Tolman theory of learning, and the effects have been studied in a variety of species and situations (**1**, **2**, **4**, **9**, **13**, **14**, **15**).

In his tentative treatment of the problem in *Principles of Behavior*, Hull (**7**) attempted to deduce shift-in-reward phenomena from an amount-of-reinforcement hypothesis. This hypothesis was based principally upon an experiment in which Gantt (**5**) conditioned a dog to four different stimuli, each reinforced by a different amount of food. These stimuli were presented in random order, and the amount of saliva evoked by each was measured at the limit of training. Hull interpreted the results (**7**, Fig. 27) as suggesting that the limit of habit strength is a positive growth function of the amount of reinforcing agent.

This hypothesis permitted Hull to predict a deterioration of performance following reduction in amount of reinforcing agent, and an increase in rate of learning following an augmentation. But the proposal had serious limitations. A group which received a small reward first, then a large reward later, could reach—but not exceed—the level of performance of a group trained with the large reward throughout, and there was associated with a given amount of reward a limit beyond which a downward-shifted group could not deteriorate.

Information on the effects of quantitative changes at variance with such restrictions was obtained by Crespi (**2**) in a study of runway performance by rats. This investigator found that the speeds of running for upward-shifted groups exceeded that for a constant reward group, and that the speeds of running for downward-shifted groups were significantly less. These results were analogous to qualitative shift data obtained by Elliott (**4**), who found that sunflower seed is less effective as a reward if given to rats accustomed to receiving a preferred bran mash. Using response latency as a measure of runway performance, Zeaman (**15**) studied quantitative shifts of reward and found support for Crespi's positive and negative contrast effects.

In the recent revision of his postulates (**8**), Hull no longer holds that amount of reinforcement is a determinant of habit strength. Instead, a negatively accelerated increasing monotonic incentive function (K) is said to participate with functions for drive (D), stimulus intensity (V),

[1] This experiment was conducted at the Primate Laboratory, University of Wisconsin, under the direction of Harry F. Harlow. It was supported in part by a grant from the Special Research Fund of the University of Wisconsin for 1948–1949.

delay in reinforcement (J), and habit strength ($_sH_R$) in the determination of reaction potential.

At the termination of learning $_sE_R$ is a multiplicative function of these functions, i.e.,

$$_sE_R = D \times V \times K \times J \times {}_sH_R.$$

The new K function is expressed in terms of weight of food given as reinforcement.

A critical test of the validity of the hypothesis is offered by th present experiment, which is concerned with the role played by differential rewards in determining a performance that can be regarded as a resolution of competing reaction tendencies. The tasks presented are reversed discrimination problems, and the *S*s are monkeys with such extensive practice on these problems that performance on new examples reaches a stable maximum after a single trial. This maximum has been shown to approximate the asymptote of a formerly continuous intraproblem learning curve (**10**). In these modified problems the relationship between the net reaction potentials of the two incompatible reactions is rapidly stabilized, and is such that the momentary reaction potential of the reversal response is almost invariably larger.

According to Hull's $_sE_R$ equation, it should be possible to modify this relationship by manipulating the amount of reward given before and after the reversal point. This variable had been held constant during the preliminary training of the monkeys, and single pieces of food were offered for correct responses. The same conditions are maintained in some, but not all, of the problems that make up the series studied in this investigation. In half of the problems the reward does not change with the reversal of the discrimination, but is constant at one or three pieces of food. At the reversal point in the other problems the reward is either increased from one piece to three or decreased from three pieces to one.

Problem performance is taken as a measure, and interproblem changes in error scores are used to indicate changes in the strengths of competing tendencies. Since the animals solve new problems at a stable level of performance, we may conclude that habit strength is no longer a construct of importance as a source of differential changes in reaction potentials. The new series of problems uses many different stimulus-objects, and the effects of stimulus intensity (V) and stimulus generalization are controlled by randomization. Delay in reinforcement (J) is not a relevant variable in the problem, for changes in this determinant of reaction potential are both small and unsystematic.

If weight of food is to be the only manipulated variable, then, the control of drive is of critical importance. It is, of course, impossible to offer food rewards without changing strength of drive, but reduction of the need upon which hunger is presumed to be based is a relatively sluggish process. The present experiment utilizes this property, and rapid changes in the level of reward prevent the association of differential drive reduction with the differential levels of reward. The number of trials per day are few, and the levels and amounts of food given are but fractions of the ad libitum consumption. A small amount of food is given each day before testing begins to inhibit the hunger contraction component of the drive state.

Since all of the proposed determinants of reaction potential are

specified from the beginning of the experiment, interproblem trends in performance under these conditions permit us to ascertain the acceptability of the *K*-function hypothesis. Hull must predict invariant performances for each of the four reward conditions.

The Experiment

Subjects.—Eight rhesus monkeys (Nos. 137, 140, 142, 143, 146, 150, and 151) were *S*s. They had previously participated in studies of discrimination learning (**11**), discrimination reversal learning (**10**), and delayed response (**12**). In the reversal learning experiment it was shown that wide variations in maintenance-food deprivation were unrelated to the development of the reversal learning set, and in the delayed response experiment it was demonstrated that the levels of food reward employed in the present study were discriminable and differentially effective.

Experimental design.—Each of the problems presented involved four reversed discriminations. The *S* was first trained to choose one member of a pair of multidimensional stimulus-objects (e.g., rubber heel, matchbox) until a criterion of performance was reached. Then the other object was rewarded until, again, a criterion of performance was attained. Retraining of the original response followed, and the series continued until the A, B, A, B, and A members had been chosen. The objects were then discarded and another pair from the laboratory's large collection substituted.

Four different conditions prevailed both before and after reversals. These were made up of combinations of low and high performance criteria with small and large rewards. The criteria were two and four successive correct choices, and the small and large rewards were one and three pieces of a preferred test food (raisins or peanuts).

The procedure differed from that of the original reversal learning experiment. Formerly, all problems involved single reversed discriminations: the animals were trained to pick first one, then the other object of a given pair.

The *S*s solved two blocks of 32 problems in the course of the experiment. The analytical datum was the number of errors made by the group in twice reaching criterion performance for a given combination of conditions. In the completely orthogonal, six-factorial experimentat design the set of main effects included learning, shift position, anteshift reward, anteshift criterion, postshift reward, and postshift criterion. Combinations of these formed a system of 49 interactions.

The reward-criterion conditions for each reversed discrimination were determined in advance. Sixteen different combinations were possible, and blocks of problems were arranged which contained all possibilities at each shift position. The procedure followed in constructing a typical block began with the random selection of a combination for the first shift, first problem. Then, since the postshift conditions for the first shift were the anteshift conditions for the second, a postshift condition remained to be chosen. In this way combinations were assigned to the first problem, and construction of the second followed. One of the remaining 15 combinations for the first shift position was chosen at random, and the cycle was repeated. The procedure yielded blocks with fairly unsystematic arrangements, and each *S* ran four such blocks.

Apparatus and procedure.—The Wisconsin General Test Apparatus (**6**, Fig. 1) was used throughout. It consists of a barred restraining cage and an adjacent table. A superstructure mounted on the table supports movable opaque and one-way vision screens, which can be manipulated by *E* from a position opposite to the cage. On the table-top is a movable test tray containing two small food-wells.

The animal was placed in the cage and given a few pieces of food. The opaque screen was then lowered in front of the cage, and one of the two food-wells was loaded with one or three pieces of food. Both wells were then covered by stimulus-objects, with the appropriate object over the well containing the food. The tray was subsequently moved forward, the one-way vision screen lowered in front of *E*, and the opaque screen raised. The monkey reached through the bars of the cage to make a choice, which was defined as displacement of either object far enough to uncover any portion of the underlying food-well. Three problems were completed each day.

Results

In the tests of significance by the analysis of variance, the fifth-order interaction has been used for error.[2] The design does not permit the testing

[2] For complete table of analysis of variance order Document 3249 from American Documentation Institute, 1719 N Street, N. W., Washington 6, D. C., remitting $1.00 for 35 mm. microfilm or $1.00 for 6 x 8 in. photocopies.

of involved main effects with the mean squares for significant first-order interactions, but this restriction does not seriously interfere with the essential interpretations. Sources of variance which were found to be significant at the .001 level are shift position, anteshift reward, postshift reward, postshift criterion, postshift reward × postshift criterion, learning × postshift reward, anteshift reward × postshift criterion, anteshift reward × postshift reward. Learning × postshift criterion is significant at the $.01 > P > .001$ level. None of the higher-order interactions met the .01 criterion.[2]

Over-all performance was remarkably stable. For the first half of the experiment the total error score was 1,402, for the second it was 1,404. The mean number of errors per shift on successive blocks of 16 problems was 1.37, 1.37, 1.36, and 1.38, respectively. From the invariance of the mean it is evident that over-all changes associated with a given level of variable are balanced by opposite, over-all changes where the other level is present.

The first reversed discrimination of a given problem is more efficient than the latter three. The errors-per-shift scores for the first to fourth shifts were 1.23, 1.43, 1.42, and 1.42. The explanation would seem to follow readily from the fact that the *S*s were trained originally in single reversal problems, but the failure of the relationship to change with practice makes this suggestion difficult to accept. The shift position variable, despite 22 opportunities, does not participate in a single significant interaction. The change from the single to the multiple reversal procedure does not affect the interpretation of other results.

Table I summarizes errors-per-shift scores as a function of learning, anteshift reward, postshift reward, and postshift criterion. All remaining significant results of the analysis of variance and the borderline interactions can be readily interpreted with these data. If scores are added to give learning × postshift reward, learning × postshift criterion, and learning × anteshift reward curves, the trends strongly suggest a common point of origin at the over-all mean. Neither reward nor criterion effects are present at the beginning of the experiment, and the two develop simultaneously. During the greater part of the experiment anteshift reward, postshift reward, and possibly anteshift reward × postshift reward effects show a marked dependence upon postshift criterion. Once readjustments are complete, however, only the interaction with postshift reward is still important. Anteshift criterion is not a significant variable, but it participates in an interesting doubtful interaction. Where the anteshift reward is large and the anteshift criterion high, postshift performance is *more* efficient than where the anteshift reward is large and the

TABLE I

ERRORS PER SHIFT AS A FUNCTION OF LEARNING, ANTESHIFT REWARD, POSTSHIFT REWARD, AND POSTSHIFT CRITERION

Postshift Criterion	Anteshift Reward	Postshift Reward	Successive Blocks of Problems			
			1	2	3	4
Low	3	1	1.55	1.48	1.53	1.42
Low	1	1	1.20	1.12	1.23	1.19
Low	3	3	1.25	1.19	1.14	1.11
Low	1	3	1.19	1.08	0.92	0.75
High	3	1	1.92	2.36	2.25	2.39
High	1	1	1.32	1.40	1.47	1.94
High	3	3	1.31	1.37	1.33	1.31
High	1	3	1.20	0.93	0.97	0.98

anteshift criterion low. The direction of this effect supports the suggestion that differential drive reduction has been controlled.

Reversal performances for each combination of levels of anteshift and postshift reward are plotted as a function of learning, in Fig. 1. The data are based primarily upon blocks of 16 problems, but because of the rapidity of changes early in the series, the data from the first eight problems are also included. The trends that have been approximated to the data meet the requirement that summations over the two levels of any variable diverge from the over-all mean by equal amounts. Thus, the course of an individual curve is determined in part by its own five points and in part by the curves drawn through the other sets of points. All have been extended to a common point of origin.

The manifestly different forms of these curves reflect, in part, an interaction between anteshift and postshift reward that is particularly indicative of the fact that given amounts of reward do not have invariant facilitative characteristics. The interdependence, however, is a transitory phenomenon. It develops rapidly at the beginning of the experiment, goes through a maximum, and then becomes progressively less evident. By the end of the experiment the interaction has nearly disappeared, and the discrepancy between the 3-1 and 3-3 curves is approximately equal to that between the 1-1 and 1-3 curves. Similarly, the 3-1 error score is as much above that for 1-1 as 3-3 is above 1-3. This change with learning is not statistically demonstrable in two-block comparisons.

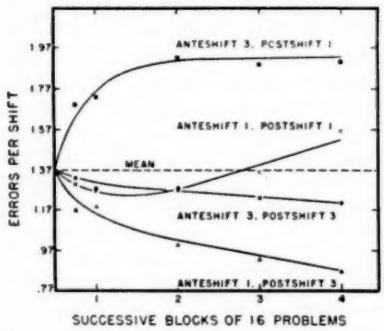

FIG. 1. Interproblem reversal performances for each combination of anteshift and postshift reward

DISCUSSION

If our conclusions were based on final performances alone, the data of this experiment would support the principal implications of Hull's *K*-function hypothesis. After readjustments are complete, varying amounts of anteshift and postshift reward have independent effects. Yet the changes which take place prior to the stabilization of the interproblem curves invalidate every assumption that is involved in postulating that a given weight of food has fixed facilitative properties. The independent quantitative manipulation of food reinforcement does not produce immediate changes in performance. Instead of the maximum effect that the Hull theory requires, initial performances reveal no changes at all. The interproblem trends in performance not only fail to parallel one another, but stem from a common point of origin.

The uniform initial performances are a most significant finding, for hypotheses concerning the means by which the varying amounts of reward acquire their differential facilitative properties are restricted. It cannot be assumed, for example, that effectiveness is uniquely dependent upon historical associations—direct or in-

direct—with varying amounts of drive reduction. The same levels of reward previously maintained substantially different delayed response performances (12), but the differences in reversal performances are gradual in their appearance. These data indicate that some learned differences in the effectiveness of varying amounts of rewards are specific to a particular behavioral situation, and that the facilitative properties of an individual level are not anchored to any process that can guarantee invariance and universal applicability.

Nor can the development of differences within the present experiment be attributed to new associations with differential drive reduction. The identity of the initial scores, and the direction of the interaction between anteshift reward and criterion, support the conclusion that controls were adequate for the elimination of this factor. The inference seems unavoidable that these differences have been learned without reference to fundamental changes in the internal organismic economy.

The changes, then, which do take place can only be attributed to the presence of a new and quantitatively superior reward, and it is apparent that a major determinant of the facilitative characteristics of a given amount of reward is the range of rewards offered for the performance of a given task. Until *S* has had ample opportunities to learn this range, it cannot be a completely operative factor, which accounts for the relatively slow development of the maximum changes in effectiveness.

During this learning period contrast shift phenomena dramatically demonstrate the influence of one amount of reward upon the other. But the data suggest that these momentary changes in the facilitative characteristics of rewards are not fundamentally different from the gradual and persistent modifications resulting from other similar experiences. The interaction between anteshift and postshift reward is marked during the period of most rapid change, but declines in importance with the stabilization of the interproblem trends. Once readjustments are complete, the facilitative characteristics of a given amount of reward no longer depend upon these individual comparisons.

It is almost certain that the stable functions obtained by Gantt (5) and by Crespi (3) are of limited generality, for the present data suggest that there is no single characteristic amount-of-reward function for a given situation. The existence of families of curves for different ranges of reward is indicated, and this proposal can readily account for the fact that asymptotes have been reported that are associated with relatively small amounts of food. The Gantt data have been criticized on this basis by Crespi (3), who points out that the maximum reward of the range presented was but a small part of the amount of food required to maintain the dog for the day.

But the reliability or representativeness of Gantt's function need not be questioned. It may well be the typical curve for the range of rewards involved, and it is not necessary to hold that the limit found for a given range of rewards establishes the limit for all ranges. Only the postulation of a simple relationship between the effectiveness of a given amount of reward and its capacity to reduce a deviation from physiological optimum makes this assumption necessary.

Summary and Conclusions

1. This experiment was concerned with the determinants of the effective-

ness of varying amounts of food reward.

2. Each of eight rhesus monkeys, previously trained in the learning of reversed discriminations, learned a series of 64 problems in which each problem involved four reversals of response to a single pair of stimulus-objects.

3. Four different reward conditions were studied: (*a*) large reward before reversal, small reward after, (*b*) large reward before reversal, large reward after, (*c*) small reward before reversal, large reward after, and (*d*) small reward before reversal, small reward after. The last condition was a continuation of the relationships prevailing during the preliminary training.

4. Predictions from Hull's *K*-function hypothesis were not supported. Instead, it was found that rewards which were differentially effective in an earlier experiment maintained initially identical performances, but that differences appeared with successive presentations of both levels of reward. These changes were shown to take place in the absence of opportunities for differential associations of levels of reward with differential amounts of drive reduction.

5. Postulations of simple relationships between the effectiveness of varying amounts of reward and their capacities to reduce physiologically defined needs are not supported. The range of rewards presented for the performance of a given task is a major determinant of the effectiveness of individual rewards.

(Manuscript received April 7, 1950)

References

1. Cowles, J. T., and Nissen, H. W. Reward expectancy in delayed responses of chimpanzees. *J. comp. Psychol.*, 1937, **24**, 345–358.
2. Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.
3. Crespi, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, **51**, 341–357.
4. Elliott, M. H. The effect of change of reward on the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1938, **4**, 19–30.
5. Gantt, W. H. The nervous secretion of saliva: The relation of the conditioned reflex to the intensity of the unconditioned stimulus. *Amer. J. Physiol.*, 1938, **123**, 74.
6. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
7. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
8. Hull, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, **57**, 173–180.
9. Lorge, I., and Sells, S. B. Representative factors in the rat under "changed incentive technique." *J. genet. Psychol.*, 1936, **49**, 479–480.
10. Meyer, D. R. Food deprivation and discrimination reversal learning by monkeys. *J. exp. Psychol.*, 1951, **41**, 10–16.
11. Meyer, D. R. Intraproblem-interproblem relationships in learning by monkeys. *J. comp. physiol. Psychol.*, 1951, **44**, 162–167.
12. Meyer, D. R., and Harlow, H. F. Effects of multiple variables on delayed response performance by monkeys. *J. genet. Psychol.*, 1951, **44** (in press).
13. Nissen, H. W., and Elder, J. H. The influence of amount of incentive on delayed response performance of chimpanzee. *J. genet. Psychol.*, 1935, **47**, 49–72.
14. Tinklepaugh, O. L. An experimental study of representative factors in monkeys. *J. comp. Psychol.*, 1928, **8**, 197–236.
15. Zeaman, D. Response latency as a function of amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

THE EFFECT OF PRACTICE WITH BRIEF-EXPOSURE TECHNIQUES UPON CENTRAL AND PERIPHERAL VISUAL ACUITY AND A SEARCH FOR A BRIEF TEST OF PERIPHERAL ACUITY [1]

BY ROBERT H. BRUCE

University of Wyoming

AND

FRANK N. LOW

Louisiana State University School of Medicine

There has been considerable interest in the problem of central and peripheral visual acuity for many years. How practice can affect these functions is not only significant for visual and perceptual theory but became of immediate and practical interest during the war. Renshaw (13), Gibson (4), Grindley (5), and Low (6, 7, 8, 9) have recently provided data pertinent to the problem.

This study attempts to furnish data on the following specific problems: (1) What is the effect of practice with brief-exposure techniques on (*a*) central visual acuity and (*b*) peripheral acuity? (2) Is it possible to select certain points in testing peripheral visual acuity that will correlate highly with a more complete test?

[1] The data for this study were collected at the U.S. Navy Pre-Flight School, Chapel Hill, North Carolina, while one of us was in the Naval Reserve and the other teaching at the Medical School of the University of North Carolina. The opinions or assertions contained herein are the private ones of the writers and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large. We are indebted to the officers and cadets of the U.S. Navy Pre-Flight School whose cooperation made this study possible, particularly Commander H. L. Hamilton. We are also indebted to the students of the University of North Carolina Medical School who served as subjects in certain portions of the experiment. A preliminary report of a portion of these data was presented at the 1947 meeting of the Rocky Mountain Branch of the APA.

Procedure

Central visual acuity was measured by *S*'s responses to Landolt broken circles having breaks of the following sizes: 2, 1½, 1, and ½ mm. The over-all measurements of the Landolt circles were five times the number assigned to them, which was the width of the break. The circles were printed in black on targets fitted to a carrier. The front card on which the circle appeared was a white surface 300 mm. square. The carrier was brilliantly illuminated in a dark room. Each *S* was seated 6 m. from the carrier. He held a black paddle in front of both eyes until a signal was given to look. The signal was given with the carrier steady. The *S* identified the position, using both eyes, and covered his eyes again. Six identifications were required for each test object in order of decreasing size. The score was the number of true answers. The *S*s had passed the rigid visual tests required of Naval aviators prior to the acquisition of these data. All the *S*s were males 18 to 27 years of age with normal or better central vision, normal color vision, and no astigmatism greater than ¼ diopter. In general the responses were prompt, the test taking about 6 min. per *S*. No attempt was made to hurry the *S*s, but they were required to make a judgment even if in doubt.

The central test was given individually to a group of cadets. The members of this experimental group were selected by picking every tenth name from battalion rosters. There were 113 cadets on whom data were compiled. These cadets as part of their regular pre-flight training were given training in the recognition of air and surface craft five days a week, 40 min. a day, in an attempt to develop efficient perceptual habits. Occasionally individuals of the group were given an additional 80-min. period on Saturday mornings or in an evening practice session. The training method devised by Renshaw (13), using spaced short and long visual exposures, furnished

an opportunity for the learning process to operate. In this method, during the training series each cadet made a judgment concerning the identification of an airplane or surface craft when the object to be identified was shown briefly, e.g., for 1/10, 1/25, or 1/50 sec. The photograph of the object was on a slide which was projected upon a screen, the length of exposure being controlled by the speed of the shutter. After this brief exposure and the judgment written by the cadet, the slide was reshown for a longer period of time. The cadets were well motivated. Competition was emphasized in all portions of the program, and the cadets were well selected. In addition, the brief and long technique provided a learning opportunity that in itself proved to predispose the learners toward those attitudes that engender superior performance.

The cadets were trained in groups of about 50. The same *Ss* were retested individually eight weeks after the perceptual training with the Landolt technique indicated earlier in this section.

A control group of 30 medical students who met the standards of vision of the cadets were similarly tested and retested in the School of Medicine. They received no visual training comparable to the cadets in the interval between tests.

Testing of peripheral visual acuity by the technique used here has previously been described in detail elsewhere by Low (**8**). Briefly, Landolt circles were employed with a modified perimeter at the following points on each eye: out 30°, up 30°, down and out 60°, up and out 30°, and out 90°. Berger (**1, 2**) has criticized the use of the Landolt circle, and Whitmer (**15**) has indicated that in testing peripheral form and pattern discrimination, the rank of the forms in percentage of accuracy in discrimination was triangle, diamond, square, rectangle, circle, and hexagon, the progression being from greater to less accuracy. Certainly the normal forces determining perception operate peripherally as well as centrally; indeed those forces such as "closure" may function in a higher degree as the image approaches the periphery.

Results

Table I indicates the scores upon the central test made on the first test (subscript 1) and the second test (subscript 2). Only at the 1- and ½-mm. target sizes are the differences clearly significant for the cadets who had been receiving the perceptual training. None of the retest scores of the medical students indicate marked improvement, although the slight improvement shown may indicate the effect of the first test given eight weeks before.

Clearly and without question there has been demonstrated an improvement in central visual acuity on the part of the cadets who had received the perceptual training as a portion of their routine. One might ask why this improvement was not demonstrated on the two larger targets. It will be remembered that both groups of *Ss* made six judgments on each size of target. The two larger targets, with these *Ss*, were rarely subject to error. This can be observed in Table I. The only function that these larger targets apparently performed was to acquaint the *Ss* of both groups with the more difficult problem posed by the smaller targets.

The next question concerned the effect of practice upon peripheral

TABLE I

Frequency of Correct Responses in Central Visual Acuity Test (Landolt Rings) with (Cadets) and without (Medical Students) Intervening Training

Size of Target (mm.)	M_1	M_2	σ_1	σ_2	t
		Cadets ($N = 113$)			
2	5.92	5.99	0.31	0.10	2.29
1½	5.86	5.90	0.42	0.36	0.84
1	4.61	5.14	1.44	1.20	4.03*
½	1.73	2.31	1.12	1.35	4.14*
		Medical Students ($N = 30$)			
2	5.97	5.97	0.17	0.17	0.00
1½	5.77	5.83	0.42	0.46	0.58
1	4.93	5.17	0.96	1.00	0.91
½	1.63	1.83	1.14	1.37	0.86

* Significant at the .01 level of confidence.

TABLE II

FREQUENCY OF CORRECT RESPONSES IN PERIPHERAL VISUAL ACUITY TEST BEFORE AND AFTER TRAINING (CADETS, $N = 113$)

Eye	Position	Size of Target (mm.)	M_1	M_2	σ_1	σ_2	t
Right	Out 30°	2	5.5	5.637	0.872	0.64	1.31
Right	Up 30°	2	2.611	3.088	1.41	1.246	3.201*
Right	Down and Out 60°	4	2.761	2.982	1.536	1.506	1.221
Right	Up and Out 30°	2	3.115	3.593	1.486	1.467	2.43
Right	Out 90°	10	2.15	2.407	1.278	1.461	1.702
Left	Out 30°	2	5.38	5.398	0.858	0.806	0.158
Left	Up 30°	2	3.035	2.938	1.482	1.435	0.602
Left	Down and Out 60°	4	2.752	2.938	1.497	1.548	1.155
Left	Up and Out 30°	2	3.327	3.257	1.374	1.441	0.482
Left	Out 90°	10	2.23	2.327	1.452	1.418	0.505

* Significant at the .01 level of confidence.

acuity. These data are given in Table II. It will be remembered that the cadets had no specific training in the development of perceptual skills utilizing peripheral stimulation, although, of course, if a cadet was not looking at the screen at the time the image was flashed upon it some peripheral training might have taken place inadvertently. While eight out of the ten scores are higher on the second test than on the first test, in only one case (the right eye—position up 30°), do we find any improvement indicating statistical certainty. The right eye, again, with the position up and out 30° has the next highest t-score. Generally speaking, the t-scores are higher for the right eye than for the left. We are inclined to believe that these data do not clearly indicate that the peripheral discriminations showed the effect of the central practice training, as was indicated in the central functions.

The next problem concerned the development of a test for central and peripheral acuity. A series of 12 coefficients of correlation was run upon both the medical students and the cadets. These correlations indicate that the scores obtained from the two smaller targets correlate highly with the total score obtained on the test, while the scores on the larger targets are not determining scores in the test. This corroborates the results of Table I.

Another series of correlations, 60 in number, was run to see if any one point tested on the periphery of the retina could satisfactorily predict the total score on the periphery of the retina. The results were negative. Nor does the index of change, either for one eye or for both eyes, indicate that such a measure can satisfactorily predict the results of the complete test. Within the limits of the points tested in this experiment each point made its contribution to the total score.[2]

DISCUSSION

To turn to the problem of the improvement of central visual acuity by

[2] We are indebted to Professor Robert J. Wherry for advice concerning the statistical treatment of the data. These correlations have been filed in tabular form with the American Documentation Institute, Science Service Building, 1719 N Street, N.W., Washington 6, D.C., and may be obtained by ordering Document 3232, remitting $1.00 for 35 mm. microfilm or $1.00 for 6 × 8 in. photocopies.

the use of short and long visual exposures, these data confirm the results of McFadden (**10**) that the level of performed acuity may be substantially raised through practice. In addition, our data indicate that group training techniques are effective as well as individual training. In reading such a review as that of Prentice (**11**) of Gibson and his colleagues' work (**4**) with these same problems, one might assume that Gibson's data indicated that individuals trained by the use of brief as well as long exposure techniques did not learn. Actually, the primary assumption involved was that the identification of air or surface craft was a perceptual process that is docile, i.e., amenable to the learning process. Gibson's data corroborate this fundamental assumption, as do the records of thousands of individuals who were trained.

Renshaw (**13**) has demonstrated that there are increases in speed and comprehension in silent reading from tachistoscopic training. Volkmann (**14**) and Wilcox (**16**) have demonstrated that visual acuity is susceptible to training. Wilcox emphasized the improvement as a function of the observer's finding and utilizing finer and finer criteria of separation of the parts of the pattern. McFadden (**10**) has emphasized the motor aspects of the transformation of the behavior act, thus placing the development of perception as one aspect of a developing motor skill. We suggest that our data are in line with this interpretation. The recent monograph of Douglas (**3**) describes the course of the development of perception when tachistoscopic techniques enable the psychologist to observe the learning process in a more detailed fashion than in our experiment. Just as Renshaw and Schwarzbek (**12**) found movement transformations, or the manner in which the act was performed, to be the most spectacular characteristic of the learning of pursuit movements, so it was possible to observe the same type of behavior in the development of this perceptual skill. When the cadets initiated training, their reactions were tentative, unsure, and they reacted to many phases of the situation that were not significant, e.g., how close a given individual was to the screen, how far off-center a particular seat was, what their particular activity had been previous to the visual training, etc. As they became skilled, the behavior involved in seeing became a skilled art in which the performers had confidence. There was less strain, less observable effort, and, in general, those psychological characteristics that are associated with the acquisition of skill became apparent.

Admitting the cadets were beginning to have some of the characteristics of skilled observers after eight weeks of practice, why do we find so little evidence that the peripheral functions have improved? Renshaw (**13**) has demonstrated that form fields show a conspicuous enlargement after tachistoscopic training. Low (**8**) and others have demonstrated that peripheral visual performance is amenable to the learning process. It seems probable that the cadets had not reached the level of skill whereby proficiency, based upon central training, could be demonstrated in a situation utilizing peripheral stimulation. In only one position (see Table II), "right up 30°" was the *t* of a size to indicate significance at the .01 level of confidence. The position "up and out 30°" was the next position indicating increased performance, again a measure on the right eye. Since the cadets were seated during

their training series and the screen was above and in front of them, it would seem plausible that the training actually was in progress. But if this suggestion were true, we would expect the same tendency to appear in measurements of the left eye. Such is not the case. Clearly, while visual central training may improve performance that is immediately a result of peripheral stimulation, if visual peripheral perceptual acuity is the primary object of the training, then practice should be initiated and continued using peripheral stimulation.

One possibility that must be faced is that the improvement in central acuity might be due to factors other than the perceptual training. The cadets' normal routine incorporated a rigid physical program. Unquestionably the cadets were in better physical trim at the conclusion of the preflight training than they had been prior to training. The medical students, on the other hand, while subject to rigorous verbal training, were not subject to such a rigorous physical routine. While visual performance might well reflect general physical condition, it seems unlikely that the central perceptual functions would reflect this general physical training more than peripheral functions.

Summary and Conclusions

This study reported the results of testing the central acuity of 113 cadets and 30 medical students. Both groups were retested after eight weeks. The cadets during the eight weeks had practice on the identification and recognition of airplanes and surface craft by a combination of spaced brief and long exposures. The cadets demonstrated a significant increase in their ability to perceive the test objects (Landolt broken circles) after the intervening practice sessions, thus showing an improvement in general perceptual ability. The medical students did not show this improvement.

One hundred and sixteen cadets were tested on a peripheral visual acuity test and then retested. No clear evidence was obtained that the training in the recognition of complex visual forms, presented predominantly to central visual areas, significantly improved peripheral visual acuity.

Coefficients of correlation analyzing the various positions on the periphery of the retina tested did not indicate that any one position tested could well predict the total score of the peripheral test.

The development of visual acuity is discussed as an example of perceptual motor skill, amenable to the learning process, rooted in the effector processes basic to an organism in the process of adjusting.

(Manuscript received April 17, 1950)

References

1. Berger, C. A comparison, under different degrees of brightness, of minimum visual distances of two luminous points and of a broken circle. *Amer. J. Psychol.*, 1942, **55**, 354–370.
2. Berger, C. The problem of peripheral visual acuity. *Amer. J. Psychol.*, 1943, **56**, 603–604.
3. Douglas, A. G. A tachistoscopic study of the order of emergence in the process of perception. *Psychol. Monogr.*, 1947, **61**, No. 6. Pp. 133.
4. Gibson, J. J. (Ed.) *Motion picture testing and research.* AAF Aviation Psychol. Res. Rep. No. 7. Washington, D.C.: U.S. Govt. Printing Office, 1947.
5. Grindley, G. C. *Psychological factors in peripheral vision.* Medical Research Council, Spec. Rep. Ser. London: 1931, No. 163.
6. Low, F. N. The peripheral visual acuity of 100 subjects. *Amer. J. Physiol.*, 1943, **140**, 83–88.

7. Low, F. N. The peripheral visual acuity of 100 subjects under scotopic conditions. *Amer. J. Physiol.*, 1946, **146**, 21–25.
8. Low, F. N. Effect of training on acuity of peripheral vision. Civil Aeronautics Administration, Division of Research, Report No. 68. Washington, D.C.: 1946.
9. Low, F. N. Some characteristics of peripheral visual performance. *Amer. J. Physiol.*, 1946, **146**, 573–584.
10. McFadden, H. B. *Visual acuity.* Duncan, Okla.: Optometric Extension Program, 1940.
11. Prentice, W. C. H. Review of J. J. Gibson (Ed.), *Motion picture testing and research. Psychol. Bull.*, 1949, **46**, 511–512.
12. Renshaw, S., and Schwarzbek, W. C. The dependence of the form of the pursuit meter learning function on the length of the inter-practice rests: II. Theoretical. *J. gen. Psychol.*, 1938, **18**, 17–29.
13. Renshaw, S. The visual perception and reproduction of forms by tachistoscopic methods. *J. Psychol.*, 1945, **20**, 217–232.
14. Volkmann, A. W. *Physiologische Untersuchungen im Gebiete der Optik.* Leipsig: Breitkopf und Hartel, 1863.
15. Whitmer, C. A. Peripheral form and pattern discrimination. *Univ. Pitt. Bull.*, 1931, **7**, 238–244.
16. Wilcox, W. W. An interpretation of the relation between visual acuity and light intensity. *J. gen. Psychol.*, 1936, **15**, 405–435.

PROBABILITY LEARNING OF PERCEPTUAL CUES IN THE ESTABLISHMENT OF A WEIGHT ILLUSION

BY EGON BRUNSWIK

University of California

AND

HANS HERMA

New York University

The trustworthiness of the sensory cues used by the perceptual system in establishing orientation in the environment is in general of a strikingly limited nature (1, 5). The dependability of cues may be defined as the statistical relationship between two environmental variables, as an "ecological validity" (5, 6). The intuitive utilization of at least some cues of an ecological validity of possibly as low as .1 to .3 by the perceptual system seems suggested by some recent studies in social perception (4, 6, 21). If the study of the acquisition of perceptual cues is to be "representative" of natural conditions, the experimental design will have to duplicate the low dependability of the ecological relationships involved (see 6, chap. X). This can be done by setting up an artificial miniature environment in which the relationship between cue and referent variable is probabilistic rather than absolute, that is, in which reinforcement is "partial" (13) rather than of the conventional unequivocal kind.

The present paper describes an early systematic effort in this direction. The experiments were conducted 14 years ago at the University of Vienna.[1] In the meantime the study was repeated, with improvements, by Levin (14, 15). In certain respects, these studies parallel in the field of perception an investigation of probability learning in rats conducted at the University of California in the spring of 1936, and published three years later (3).

[1] A brief report of the study was given by Brunswik at the International Congress of Psychology in 1937 (2).

Experimental Design

Medium.—The response used in our study as a test of perceptual learning was the well-known so-called "weight expectancy illusion" described by Müller and Schumann (16), Steffens (18), and Myers (17). In this illusion, lifting of a relatively heavy weight is found to elicit an underestimation of weight in subsequent trials, and vice versa for light weights. The phenomenon may also be described as a successive weight contrast. Its high sensitivity made it seem the ideal medium for our study. Under this illusion, there is a dual determination of each weight perception, one by the anticipatory cue and one by the subsequent direct impact.

In our case, the problem was restated as one of differential response in simultaneous lifting of pairs, and a probabilistic design was employed to induce the contrast effect. The major cue-variable was defined by two discrete lateral positions, presentation to right vs. left hand. In order to reinforce the position cue in immediate perception, yet at the same time to introduce an additional factor that may prove confusing to explicit penetration of the design by the *S*, objects were made in two colors. For each *S*, there was a fixed relationship of position (hand) and color, that is, the two cue-variables were confounded with each other, or artificially "tied" (6), except for two special experimental series not carried beyond the exploratory stage; for the sake of brevity, we will describe our experiments in terms of position throughout this report. The significate or referent variable, weight, was also handled in a dichotomous manner by being limited to two discrete and

markedly different levels of strength, to be called "heavy" and "light."

Objects.—The objects used in lifting were of three types. Objects in group I consisted of wooden blocks of 10 cm. height with a square base of 4 cm. by 4 cm. Since it seemed possible that the excess finger pressure required in lifting the blocks could interfere with the contrast effect, objects in groups II and III were given the shape of a flower pot 6 cm. in height, 4 cm. at the base, and 7 cm. at the top, that could be suspended from the finger-tips. The objects were drilled and some of them evenly filled with lead to adjust for weight. They were lacquered to prevent identification by visual inhomogeneities. The standard "light" and "heavy" weights were 100 and 200 gm. for the blocks and 100 and 500 gm. for the pots, with duplicate sets in red and green.

In order to loosen up the *Ss*' impressions of equality in the test trials, the green standards in groups I and II were replaced in the test trials by two "approximates" slightly above and below the respective standard and alternating with each other from one test trial to the next to neutralize the deviations over the series. The deviations were 10 per cent up and down (in terms of the lesser weight) in group I, and 20 per cent in group II. For group III, only standard weights were used.

Experimental series.—Over-all relative frequencies of position-weight combinations were made complementary to each other on the two sides, but their sequences were kept independent of each other and were quasi-random. In our training sequences each trial is defined by a quasi-incidental pairing of two individual position-weight combinations. If on both sides the majority trend of reinforcement for the respective side happens to prevail, we speak of a positive (+) pair. The relatively rare pairings of minority combinations constitute negative (−) trials. Pairings of a majority with a minority combination will be called balanced (0); in this case, presentations on the two sides are either both "heavy" or both "light." These units of the training sequence offer a sensitive medium for whatever weight illusion may be developing and thus were chosen to fulfill the added function of test trials.

In all, there were eight experimental series. Since in the above mentioned experiments with rats (**3**) a reinforcement ratio of 2 to 1 (67 per cent) had been found to be just below the threshold of probability learning, while 3 to 1 (75 per cent) was distinctly above threshold, and since at the same time 2 to 1 (or $\phi = .33$, see below) is of a similar order as the ecological validities of what appear to be the most dependable bodily cues in the natural social perception of human traits, this ratio was chosen for the position-weight association of our *main experiment*. The minimum by which the over-all frequency ratios on the two sides can be realized consists of $(2 + 1) \times (2 + 1) = 9$ pairings. Of these, $2 \times 2 = 4$ are positive, $1 \times 1 = 1$ pairing is negative, and $(2 \times 1) + (1 \times 2) = 4$ are balanced. The basic cycle chosen is shown in Table I. In order to ascertain, for our group of *Ss*, the degree of approximation to complete absence of the illusion, i.e., to "chance" response, prior to the introduction of trials upsetting the hand-weight balance, we have compromised the principle of random sequence of trials to the point of having the first cycle begin with two balanced (and, thus, testing) trials. The total training series consisted of eight basic cycles, four of them in opposite order to increase random-like variety of sequence, and running into each other to form an uninterrupted training sequence of 72 trials (top of Fig. 1).

TABLE I

BASIC CYCLE OF NINE TRIALS IN THE MAIN EXPERIMENT

Trial No.	Presentations		Trial Type
	Hand A	Hand B	
1	Heavy	Heavy	Balanced (test)
2	Light	Light	Balanced (test)
3	Heavy	Light	Positive
4	Heavy	Light	Positive
5	Light	Heavy	Negative
6	Light	Light	Balanced (test)
7	Heavy	Light	Positive
8	Heavy	Heavy	Balanced (test)
9	Heavy	Light	Positive
Total	Heavy: 6 Light: 3	3 6	Positive: 4 Balanced: 4 Negative: 1

The center bottom portion of Table I represents a four-fold table. Since the two variables, position and weight, have been made dichotomous in the artificial ecological universe represented by our experimental design, the degree of intimacy of their association may be described in correlation terms by Pearson's mean square contingency coefficient for two discrete variables, known as ϕ (phi) (**9**, p. 352 f.). For the main experiment, this coefficient is .33. The accumulation of ecological evidence regarding hand-weight correlation is shown graphically in the middle part of Fig. 1. At the beginning there is gross fluctuation after each new trial, but after

the conclusion of each cycle the coefficient is always exactly .33. There is very close approximation to the over-all ratio after the second cycle.

For the full series of the main experiment 36 *Ss* were used, 12 each with object groups I, II, and III. The sequences of substitute approximate weights used with object groups I and II were opposite to each other so that the results of each individual test trial are strictly comparable with the next for the total group of 36 *Ss*.

Since the first 18 trials (two cycles) of the main experiment seemed the most crucial from the point of view of results, they were duplicated with a new set of 36 *Ss* (*short series*).

To study relearning, for group III of the short series (12 *Ss*) the original 18 trials were followed without interruption by a *reversal series* duplicating the original series with opposite sign for each non-balanced presentation pair. This is shown at the top of Fig. 2; the upper curve represents the cumulative ecological position-weight association which swings down for the second 18 trials as negative evidence accrues, ending at chance level.

A *higher reinforcement ratio*, 4 to 1 (80 per cent, $\phi = .6$), was employed in a further experiment. The minimum cycle, to which this series was limited, consists of $(4 + 1) \times (4 + 1) = 25$ trials, made up of $4 \times 4 = 16$ positively reinforcing pairs, $(4 \times 1) + (1 \times 4) = 8$ balanced pairs, and $1 \times 1 = 1$ pair of negative unbalance. Their quasi-random sequence (with the negative trial near the middle) was as follows:

$$00+++0++0++++$$
$$-+0++0++++00.$$

There were 12 *Ss*. Object group II was used; to offset the possible effect of the use of approximate weights, results of adjacent test trials were pooled in four pairs.

Conventional training, that is, perfect ecological correlation (100 per cent, $\phi = 1.00$) was studied by means of a straight series of 18 positive reinforcing trials, followed by four balanced test trials to get an idea of the extinction rate for our type of response and, thus, indirectly, of the pre-experimental trend of judgment in the present series; *Ss* and weights were as in the 4 to 1 series.

The possibility of a *one-trial* establishment of our contrast illusion was likewise studied, using 12 additional *Ss*. A single initial reinforcement trial was followed by two balanced trials.

Separation of the position from the color cue was undertaken in two special series. The number of *Ss* was 12 in each of them, using object group III. But in this case part of the *Ss* had participated in one of the earlier sessions. These series are therefore no more than exploratory. In the first, three positive reinforcing trials were followed by two balanced test trials in which the colors shifted position. In the

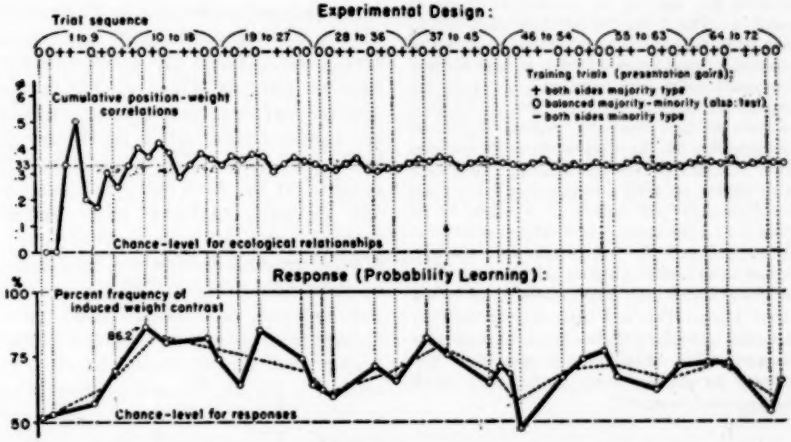

FIG. 1. Design of main experiment with an artificial ecological association of $\phi = .33$, and probability learning in 36 *Ss*. (Dotted vertical guide lines connect test trials in presentation sequence with corresponding per cent frequencies of illusion, passing between adjacent points of cumulative ecological correlation curve.)

second, there were 18 trials in which there was consistent (100 per cent) association of weight and color combined with an over-all neutrality of the position cue (50 per cent positive and 50 per cent negative position-weight reinforcements switching in quasi-random sequence). This was followed by four balanced test trials, with a continued switching of the position of the colors.

Subjects, procedure, instructions.—The *S*s were psychology students or other persons with some psychological background. Except for the two exploratory special position vs. color series, each of the *S*s was allowed to participate in one experimental series only. In all experiments, for half of the *S*s the preponderance of heavy over light weights was on the right side, and for half, on the left.

Experiments were performed individually. The *S* was seated comfortably in front of a screen placed on a table. Behind the screen sat *E*, arranging the stimulus objects for presentation. These were placed on a felt-covered, sound-absorbing tray that could be slid through an aperture in the lower part of the screen so as to come within easy reach of the *S*. The *S* was told that a series of pairs of objects similar to a sample pair now before him would be presented to him for simultaneous lifting, and that after some but not all of the presentations, he would be asked which of the two objects appeared heavier at the first moment of lifting. Immediate impression was to be expressed as a snap judgment. The so-called third category, "equal" (including "undecided"), was permissible but was to be avoided as much as possible. The *S* was further instructed not to look while the objects were moved in and out of position. Upon a signal he was to turn his eyes downward so as just to be able to locate the objects without fixating them, the unexplained purpose of this being to prevent individual identification of the weights through minimal cues. He or she then was to grasp the objects with fingers pointing downward and lift them simultaneously and rapidly to a height of no more than a few inches, the right object with the right hand and the left object with the left, setting them back in place as quickly as possible, the entire lifting procedure to take no more than about 1 sec. This was then demonstrated by *E* taking the place of *S*. There were no preliminary trials. Each of the test trials was followed by the question: "At this trial, which one appeared heavier?"

Results

A case of contrast as defined in this paper is given when of the two objectively equal (or approximate) weights of a balanced trial, in its function as a test trial, the one presented on the side with generally lesser frequency of heavy objects is judged as the heavier of the pair; if it is judged lighter, we have an assimilation judgment. The over-all contrast-effect induced by the probability training is defined as the percentage of contrast judgments, including half of the equal judgments. A 50 per cent contrast-effect represents chance response. Any significant rise above chance, although indicating the establishment of a perceptual "illusion," at the same time shows that probability learning is effective; it thus represents the over-all "correct" response from a learning point of view.

Main experiment and short series.—While object groups I and especially II yielded a smaller total of equal judgments over the entire series, the over-all contrast effect as defined above is very close for the three groups (within about 2 per cent of each other). Results have thus been lumped together for the 36 *S*s on each of the 32 balanced trials serving as test trials. They are shown in the bottom part of Fig. 1. The probability learning curve rises rapidly from near-chance to a maximum of 86.2 per cent. This maximum is reached at the fifth test trial, that is, the eleventh training trial. Afterwards, there is a somewhat paradoxical gradual decline to a final value in the neighborhood of 65 per cent in the face of continuing trend reinforcement. The general trend of the major and minor peaks, located at trials 11, 22, 38, 55, and perhaps also at trials 65 to 67, is one of consistent decline. The over-all trend is best shown by the thin broken line, a smoothed curve obtained by averaging the values on the first and second,

then those on the third and fourth tests, and so forth.

Results of the short series are compared with those of the main series in Table II. There is close correspondence of the general trend for the two sets; the eleventh trial marks the peak for both of them, and there is a consistent downward trend from then on. The upward swing is statistically significant, with a critical ratio of 4.5 between trials 1 and 11 for the entire group of 72 *Ss*. The paradoxical downward trend is likewise significant, with a critical ratio of 2.7 between trials 11 and 18. The best-fitting linear slope for the entire material presented in Table II is 3.3 per cent on our scale per unit step from one of the eight test trials to the next ($r = .64$); this is significant at better than the .01 level. The non-linearity of the trend is likewise significant at better than the .01 level ($\eta = .94$).

More detailed inspection of the probability learning curve of the main experiment reveals that a definite recency-effect appears to be superimposed upon the general shape of the curve. While the over-all initial rise followed by a slow downward trend is obvious from an inspection of the smoothed curve in Fig. 1, the actually obtained values as given by the heavy solid curve present a more uneven picture. Each of the minor upswings which may be noted follows upon a sequence of positive, trend-supporting presentations which is interrupted by no more than scattered balanced trials and is free of negative trials. The deepest depressions, on the other hand, are found at the only two trials which conclude an accumulation of as many as four balanced presentations in a row (actually the most trend-weakening ecological evidence in the entire series). These trials are 29 and 47.

Reversal series.—Results for this series, smoothed in the manner described above for the thin broken curve of the main experiment to eliminate incidental factors, are shown by the lower curve in Fig. 2. During the reversal period the curve drops far below the chance level, coming closer to chance again at the very end. Both this downward and paradoxical upward trend are quite symmetrical to the up and paradoxical down of the original training.

Higher reinforcement ratio.—Per cent frequencies of weight contrast, averaged for pairs of the eight test trials of this series in a manner similar to Fig. 2, are as follows: 42, 92, 84, 83. Since the number of *Ss* is small, it may be stated only tentatively that there again is an upward and subsequent downward trend, as familiar from the main experiment, although the rise is

TABLE II

PER CENT FREQUENCY OF INDUCED CONTRAST FOR THE TWO INITIAL CYCLES OF THE MAIN SERIES AND FOR THE SHORT SERIES

Experiment	Test Trials							
	1	2	6	8	11	13	17	18
Main series (36 *Ss*)	51.4	52.8	57.0	69.5	86.2	80.5	82.0	73.6
Short series (36 *Ss*)	54.2	57.0	59.7	68.2	87.6	77.7	64.0	62.6
Totals (72 *Ss*)	52.8	54.9	58.4	68.9	86.9	79.1	73.0	68.1

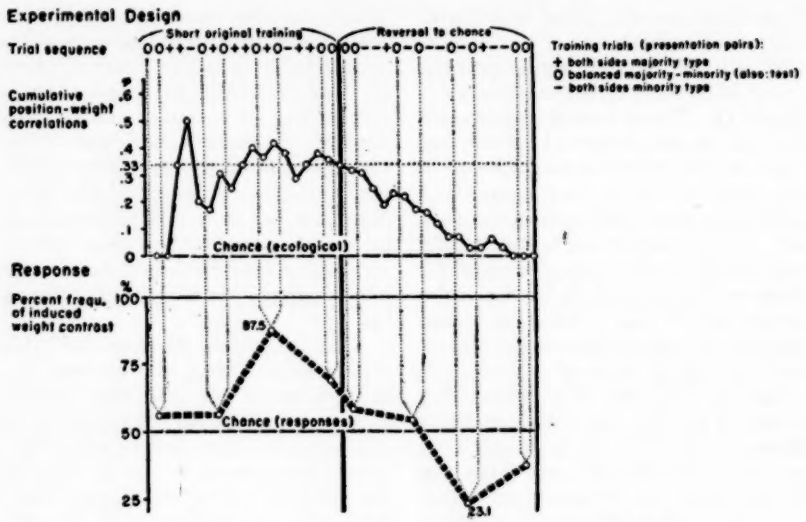

FIG. 2. Reversal to ecological chance after short series, and probability learning-relearning pattern, in 12 *Ss*

to a greater height despite the below-chance start of the curve, and the tapering-off is toward a higher level.

Conventional training.—For the first two of the four test trials following the 18 continuous positive trials of this series, the contrast effect was complete (100 per cent) for the 12 *Ss* employed; for the combined responses on the last two trials, the induced contrast dropped to 65 per cent, indicating a relatively strong effect of recency and a rapid rate of extinction.

One-trial learning.—The contrast score after the single reinforcement was 87.5 per cent for the first and 75 per cent for the second of the balanced trials, each figure representing this time, to be sure, only one trial per *S*. These results suggest quick establishment of the characteristic response, although pre-experimental position tendencies, not testable in this particular case, may, of course, have considerably deviated from the ideal chance level of 50 per cent with such a small number of *Ss*.

Exploratory position vs. color series.—In the first of these two series, the contrast score on the first test trial was perfect (100 per cent) in terms of position in spite of the reversal of color; if color rather than position had been the more effective cue, the score should have been significantly under 50 per cent. On the second test trial the score fell to 75 per cent, indicating perhaps that there had been not too much pre-experimental bias in favor of the result on the first test trial.

The second exploratory position vs. color series, in which the training was carried by color alone while position was kept neutral, had no appreciable training effect. For both the first and second pairs of test trials the score was 46 per cent while it should be significantly above 50 per cent if color were picked up as a cue for weight.

Discussion

Long-range trend of probability learning; paradoxical decline.—Successful establishment of an extrinsic cue for weight, in spite of the fact that its over-all ecological validity is no better than 2 to 1, reveals superiority of sensitivity of probability learning in this particular case of human perception over the particular case of probability discrimination in rats studied by Brunswik (3). Over and above all the more obvious differences in problem and procedure that make a fair comparison well-nigh impossible, it must be stressed that the corresponding rat experiment guaranteed exploration of both sides only on those trials on which the first turn was to the unrewarded side and the animal had to "correct" itself by turning to the opposite side. In the present experiment information is always complete and even simultaneous on both sides. Furthermore, the training period was only one-third as long in the experiment with rats as it is in the present experiment. An experiment by Jarvik (12) has shown that low reinforcement ratios may yet produce learning after a prolonged latency period.

The fact that the initial upswing of the probability learning curve is followed by a tapering-off despite continued reinforcement is quite atypical of ordinary learning curves obtained with the standard 100 per cent reinforcement design, and has, therefore, been called "paradoxical" in describing the results. Apparently, probability learning has certain characteristics of its own. It may well be that the learning mechanism as a whole is characterized by a certain tendency toward an initial all-or-none response (3). If justification of a positive response is not complete in the further course of reinforcement, there apparently follows what purposive behaviorists like Tolman (20) may be inclined to consider a disappointment or disruption of (verbalized or unverbalized) expectancies. Such a type of reaction would imply an eventual settling on a compromise level of frequency, as observed in our material. This in-between solution would be somewhat of an analogy to the well-known compromise tendencies that spoil the accuracy of the perceptual constancies (1, 6). So long as we limit our consideration to the evidence presented here, however, the possibility cannot be excluded that the paradoxical decline of our probability learning curve is wholly or in part an artifact of our use of an illusion as the test of learning; as is well known from texts in experimental psychology, some illusions have been shown to decline through "practice."

From a purely descriptive point of view of our main experiment shows a fairly close agreement between the level of the final, near-horizontal stretch of the learning curve covering approximately the last three training cycles, within which an average induced contrast-effect of about 65 per cent is maintained, and the ecological frequency ratio of 2 to 1, or 67 per cent. A similar picture is given in our experiment with a 4 to 1 ratio. We would hesitate to assume, however, that a general principle underlies this correspondence, as appears to have been claimed by some authors, notably by Hilgard (10, p. 272), in his discussion of Brunswik's experiment with rats (3) and, perhaps, also by Grant and Hake (8).

Follow-up experiments by Levin (14, 15), in which the standardized random sequence of the basic cycle of our main experiment was systematically evened out by the use of a total

of nine alternate types of cycles, each of them presented to a different group of *Ss*, further confirmed the characteristic shape of our probability learning curve with its initial upward swing followed by a gradual downward trend, although the peak was found to be somewhat delayed.[2]

In an experiment by Jarvik (**11, 12**), dealing with the differential anticipation of one of two alternatives in randomized word sequences, a reinforcement ratio of as low as 60 per cent vs. 40 per cent was found to be above the threshold of probability learning, although there was a considerable delay of the peak. There also were suggestions of ensuing decline, but they were far from being statistically significant.

Recency-effect.—The positive influence of recency in temporarily raising the learning effect was revealed by a detailed study of our main experiment. It is also evident in the dropping of the reversal training curve *below* the chance level, in the effectiveness of short-term or one-trial reinforcement, and in the relative rapidity of extinction after extensive conventional absolute reinforcement; the latter is in line with Fieandt's (**7**) finding of rapid extinction in case of artificially established illumination cues in perceptual color constancy.

Jarvik (**11, 12**) has reported a shift from the positive to a negative recency-effect during continuous "runs" of positive reinforcements in probability series. There is no analogue to this in our results, indicating that the phenomenon may be specific to certain conditions, perhaps those favoring an analytical attitude toward the single response.

[2] For a brief discussion of other aspects of Lewin's study relevant to our problem, see Brunswik (**6**, p. 54).

TABLE III

FREQUENCY OF AWARENESS OF RELATION OF CUES TO WEIGHT

S's Report	Position Cue	Color Cue
Correctly aware of cue	4	5
Incorrectly aware of cue	16	14
Undecided about cue	2	3
Total	22	22

Learning and awareness of what is being learned.—In order to throw light upon the relationship between the learning of a cue by the perceptual system and the explicit intellectual awareness of the cue on the part of *S*, 22 of the 24 *Ss* participating in those of the short series which were not followed by additional reversal training (groups I and II) were asked the following questions after completion of the experiment: "In your opinion, which objects were on the average heavier, the green or the red?" After answering this, the *Ss* were asked a similar question concerning ". . . the right or the left?" Those *Ss* who in this second case seemed uncertain as to whether they should refer to the judgments actually asked for by the *E*, or to their impressions in the total trials regardless of whether or not these impressions were overtly expressed, were told to refer to the latter.

Results are shown in Table III. Conscious impressions in terms of right vs. left position seem in four-fifths of the cases in opposition to the correct. Explicitly conceptualized learning thus seems in our case to be determined chiefly by the overtly verbalized contrast illusion predominant in the test trials with their objective balance of weights, rather than by the objectively and undoubtedly also subjectively far more drastic opposite differences in the

trials with positive (vs. negative) imbalance. Yet, the existence of our perceptual contrast effect reveals the acquisition, on a more primordial level, of a concurrent implicit quasi-expectancy in the correct direction. This case of learning with "submerged mediation" (1) is not only what Thorndike (19) and others have referred to as "learning without awareness of what is being learned"; it is learning *against* awareness of what is being learned. Restriction of verbalization to the test trials has made it possible for the two levels of learning to stand out more clearly against each other.

As is further shown in Table III, the overt effect seems to hold about equally well for the other potential cue, color (14 vs. 5 cases). In part this may be due to the fact that, as was revealed by a further question to the *Ss* (results not shown in the table), four of them had become aware of the univocal relationship between position and color, and another 15 thought that such a relationship was maintained in the majority or in most of the trials; only three *Ss* claimed that there was no relationship between color and position.

Summary

Artificially established ecological associations of 2 to 1 (67 per cent, $\phi = .33$) and of 4 to 1 (80 per cent, $\phi = .60$) between position (right- vs. left-hand presentation) as a perceptual cue, and weight as the referent variable were found to be effective in inducing an illusion of weight contrast. The probability learning curve, after first rising rather rapidly, shows a subsequent slow but steady decline to a compromise position. The decline is "paradoxical" in view of continued reinforcement of the over-all association. This decline may be an intrinsic characteristic of probability learning or merely the result of our use of an illusion as the test of learning. A short-range positive recency-effect is superimposed upon the characteristic long-range trend.

Added experiments deal with reversal of training, one-trial learning, and an exploratory establishment of the position cue, rather than a confounded color cue, as the one chiefly utilized by the perceptual system in our experiment.

Under the conditions of our experiment, perceptual probability learning seems not only not to be based on, but to run counter to, what is being learned at the conscious level.

(Manuscript received April 14, 1950)

References

1. Brunswik, E. *Wahrnehmung und Gegenstandswelt.* Vienna: Deuticke, 1934.
2. Brunswik, E. Das Induktionsprinzip in der Wahrnehmung. In H. Piéron and J. Meyerson (Ed.), *Communications, 11th International Congress of Psychology, Paris, 1937.* Paris: Alcan, 1938. P. 346. (Abstract.)
3. Brunswik, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, **25**, 175–197.
4. Brunswik, E. Perceptual characteristics of schematized human figures. *Psychol. Bull.*, 1939, **36**, 553. (Abstract.)
5. Brunswik, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255–272.
6. Brunswik, E. *Systematic and representative design of psychological experiments: with results in physical and social perception.* Berkeley: Univ. of California Press, 1947.
7. Fiandt, K. v. Dressurversuche an der Farbenwahrnehmung. *Arch. ges. Psychol.*, 1936, **96**, 467–495.
8. Grant, D. A., and Hake, H. W. Acquisition and extinction of the Humphreys' verbal response with differing percentages of "reinforcement." *Amer. Psychologist*, 1949, **4**, 226. (Abstract.)
9. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.

10. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
11. Jarvik, M. E. Probability discrimination and the gambler's fallacy in guessing. *Amer. Psychologist,* 1946, **1**, 453–454. (Abstract.)
12. Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.,* 1951, **41**, 291–297.
13. Jenkins, W. O., and Stanley, J. C. Partial reinforcement: A review and critique. *Psychol. Bull.,* 1950, **47**, 193–234.
14. Levin, M. M. Weight illusions induced by cues of low validity. *Psychol. Bull.,* 1943, **40**, 582. (Abstract.)
15. Levin, M. M. Inconsistent position cues in the establishment of perceptual illusions. Unpublished Doctor's dissertation, University of California, 1946.
16. Müller, G. E., and Schumann, E. Ueber die psychologischen Grundlagen der Vergleichung gehobener Gewichte. *Pflüg. Arch. ges. Physiol.,* 1889, **45**, 37–112.
17. Myers, C. S. *A text-book of experimental psychology.* New York: Longmans-Green, 1911.
18. Steffens, L. Ueber die motorische Einstellung. *Z. Psychol.,* 1900, **23**, 241–308.
19. Thorndike, E. L. *The fundamentals of learning.* New York: Columbia Univ. Press, 1932.
20. Tolman, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century, 1932.
21. Wallace, R. P. Apparent personality traits from photographs varied in bodily proportions. *Psychol. Bull.,* 1941, **38**, 744–745. (Abstract.)

PROBABILITY LEARNING AND A NEGATIVE RECENCY EFFECT IN THE SERIAL ANTICIPATION OF ALTERNATIVE SYMBOLS [1]

BY MURRAY E. JARVIK

University of California

This is an experiment designed to be "representative" of the establishment of cues of limited "ecological validity" (2, 3). An experiment of this kind creating a weight expectancy illusion by the use of an artificial position cue on a probability basis has previously been reported by Brunswik and Herma (4). A follow-up of this experiment has been accomplished by Levin (8, 9). Brunswik also has reported an experiment on probability discrimination with rats (1). In the present study an objective "probability" (i.e., relative frequency) of the occurrence of one as against the other of two events in a series—words or pencil marks in this case—was presented to a number of adult human *Ss*. Preceding events were to be established as cues for subsequent events in the series. The experiment was conducted in 1946, and a preliminary report was given (7). In the meantime, Grant and Hake (5) have reported a study conducted under conditions somewhat similar to those of the present experiment and paralleling it so far as the long-range analysis of probability learning with our highest reinforcement ratio (75 per cent) is concerned. Alternation trends in guessing somewhat related to our short-range negative recency effect have been described in the literature, notably by Skinner (10).

[1] The writer is indebted to Professor Egon Brunswik for advice in the preparation of this paper.

Experimental Design

The *Ss* were supplied with mechanic's rule paper, and given the following directions: "I am going to read you a series of words. They will either be 'plus' or 'check.' Before each word I am going to give the signal 'Now.' With each 'Now' I want you to guess what will follow, that is, whether the next word is going to be a 'check' or 'plus.' The series will start with a 'Now.' In response, you write either a plus-sign (+) of a check-mark (√) in the left side of the uppermost cell of your protocol, depending on what you anticipate to come. After you have given your response in this manner, I shall present to you the correct answer and you write this in the same cell immediately to the right of the mark representing your guess. With each subsequent 'Now' take the next lower cell until you reach the end of the column. Remember, the object of this is to anticipate each subsequent word as accurately as you can. Time between presentations will be 4 sec., controlled by a metronome going at the rate of one tick per second. Every four ticks I shall say 'Now,' and at the second tick between the 'Nows' I shall say either 'check' or 'plus.'"

The "checks" and "plusses" presented as the correct solutions were arranged in three randomized series, with separate groups of *Ss* for each series. There was an over-all predominance of "check" over "plus" reinforcement in each series. In Series A, the proportion of "checks" to the total was to be 3:5, or 60 per cent; in Series B, 2:3, or close to 67 per cent; and in Series C, 3:4, or 75 per cent. The series varied somewhat in length. Data are complete up to at least 87 guesses in all series; our presentation will be limited to those so as to make results for the three series comparable to each other. Each series was carried out without a stop. The over-all frequency ratios listed above were maintained within segments of from 8 to 20 presentations in various parts of the different series, the lengths of the segments being defined as multiples of the minimum blocks of 5, 3, and 4, respectively, for which the given proportion could be made to hold. The position of the "plusses" within the segments was arranged with the use of Fisher and Yates's table of random numbers (6) for each individual segment

independently of any other segment. Since the the segments followed upon each other continuously and were thus not as such recognizable to the *Ss*, homogeneous runs of up to four "plusses" and of up to six "checks" in a row have occurred incidentally in our presentation series. Furthermore, since the beginning or the end of any series did not necessarily coincide with the beginning or end of a segment, the over-all probability ratios given above hold only with approximation for the respective total series.[3]

Choice of the over-all probability ratios was determined by the fact that in experiments with rats by Brunswik (1) a ratio of 67 per cent had been found to lie below what he has termed the "threshold of probability discrimination," while the same ratio was clearly above threshold in the Brunswik and Herma experiments with perceptual cues (4). Our over-all frequency ratios move in the same general area, attempting to test maximum sensitivity.

The experiments were performed in groups. The *Ss* were students in a course in introductory psychology. There were separate, non-overlapping groups for the three series, with at least 21 *Ss* for each series. In the present report, data are utilized in full when groups are larger, as is the case for Series A ($N = 29$) and C ($N = 28$).

Results and Discussion

Long-range trend of probability learning.—Since the over-all trend in guessing is overshadowed by a great deal of fluctuation from trial to trial—largely due to a short-range negative recency effect to be discussed in the next section—learning curves must be considerably smoothed to show the over-all trend. The sequence of 87 guesses was divided into seven groups of 11 and one group (the last) of 10 trials, and the mean percentage of "check" guesses ascertained.

[3] Among the 86 presentations interspersed between the 87 guesses considered, there are 51, or 59.3 per cent, "checks" in Series A; 59, or 66.3 per cent, in Series B; and 64, or 74.4 per cent, in Series C. Since in the case of Series B the actual percentage in the slightly longer sequence considered for the preliminary report (7) was closer to 70 per cent than to the intended 66.7 per cent, the former figure was used there for designation.

Results are shown in Table I. In all series *Ss* show an initial slight bias in favor of "plus" anticipations (concerning the establishment of 50 per cent as the chance level, see 10). Sooner or later this gives way to a consistent preponderance of "check" anticipations. A linear component of incline as described by the *b*-coefficients (average increase of best-fitting straight line on our percentage scale from group to group) and by the *r*'s is statistically significant for all three series at better than the .001 level of confidence. All three presentation ratios are thus above the threshold of probability discrimination under the conditions of our experiment. Most clearly this is the case for Series C with its 75 per cent reinforcement of "checks" over "plusses" where the figures rise fastest and highest (to 80 per cent in the fourth group). By contrast, results of Series A with its mere 60 per cent reinforcement show the slowest and least pronounced rise, but the differences among Series A, B, and C are not statistically significant. Throughout the first three groups of trials in Series A there is a prolonged latency period during which the figures stay below 50 per cent, and the obtained peak is low and considerably delayed (62 per cent in the sixth group).

In further following up the trend, all three rows of data are somewhat remindful of the learning pattern found characteristic of probability discrimination by Brunswik and Herma (4) and by Levin (8, 9), with its initial upswing and subsequent "paradoxical" decline, although the trend is fluctuating and the decline not significant in our case. Whether or not the curves eventually tend to settle at the levels of the respective objective probabilities of reinforcement, as claimed by Grant and Hake

TABLE I

TREND OF "CHECK" ANTICIPATIONS FOR THE THREE PROBABILITIES OF REINFORCEMENT, WITH THE 87 GUESSES OF EACH SERIES FRACTIONATED INTO EIGHT GROUPS

Series	Per Cent "Check" Reinforcements	Mean Per Cent "Check" Anticipations in Successive Groups*								b (Slope)	r (group order × percentages)	P_r
		1	2	3	4	5	6	7	8			
A	60	45.6	45.8	46.1	53.6	61.5	62.2	55.5	57.3	2.4	.36	.001
B	67	45.9	51.8	64.1	67.5	61.1	62.6	57.9	67.4	2.7	.38	.001
C	75	45.6	56.4	66.7	80.4	79.8	70.3	77.1	75.9	4.0	.41	.001

* There are 11 trials in the first seven groups, and 10 trials in the eighth group.

(5), must be left open here. In his study with rats, Brunswik (1) had on the descriptive level already found suggestions of such a correspondence, but had refrained from generalization, and so do Brunswik and Herma (4) in their paper on induced weight contrast.

Turning to an analysis of single trials (not shown in the table), it should first be noted that on three occasions in Series C, all 28 *Ss* anticipate a "check" on the next trial, thus establishing a 100 per cent score. As is to be expected, all three are located in the second half of the training sequence, to wit, on trials 48, 64, and 83. The probability of such unanimity to occur by chance on a single trial is only .0000000037; the fact that there are 87 trials on which it may occur raises the probability for the entire series to .00000032, which still is a very small chance probability. The highest scores on single trials for the remaining two series are about 90 per cent; more specifically, there are only two "plus"-anticipations out of 21 guesses on trial 79 of Series B, and only three out of 29 on trial 61 in Series A, again establishing probability learning at a high level of statistical significance ($p = .001$, and .00007, respectively, for such majorities to occur by chance once in 87 trials).

It should be added that none of the five high scores listed can be ascribed to immediately preceding prolonged incidental runs of "check" reinforcement and thus to the positive recency effect known from classical learning theory and described for probability learning by Brunswik and Herma (4). For the one trial of Series A with a high score, as referred to above, even the opposite is true, since the high score follows upon the only continuous run of four "plus" (*not* "check") reinforcements in our entire material (see below and Fig. 2); it thus demonstrates with particular clarity the *negative* recency effect, at the same time making clear that it is the result of an over-all probability learning effect of which the negative recency effect must be considered but a part.

The negative recency effect.—We turn now to incidental serial accumulation of "check" or "plus" reinforcements in homogeneous "runs." Since there are more "check" than "plus" presentations, such runs are apt to be more numerous and longer for the former. In Fig. 1 a survey is given of results on all runs of two or more "check" reinforcements occurring in the three series. Only runs preceded or followed by a "plus" are considered. In the bottom part of the figure the number of runs for each series is shown. The curves are based on mean percentages of "check" anticipations following each "check" reinforcement of the run and including those following the immediately preceding and the subsequent "plus" reinforcement.

As we move from the left to the right in Fig. 1, curves start higher and higher. This is a direct result of the fact that extended runs of "check" reinforcements are more frequent in series with higher over-all reinforcement ratios and a more pronounced learning effect. Also, the longest runs happen to be in the later portions of the training. After the first "check" reinforcement there is a

tendency toward a rise in the curves, in conformity with the positive recency effect. After the second "check" reinforcement, however, the curves drop consistently despite continuing positive reinforcement. In other words, there is a tendency on the part of the *Ss* to expect a continuous sequence of the generally more frequent of the two alternative words to lead to a change-over to the other alternative on the next trial. This phenomenon may be characterized as a "negative recency effect."

The interference of the negative recency effect with the over-all trend of probability learning is so great after three to four "checks" for the various curves that all gains are temporarily obliterated, and after four or five "checks" the preponderance of anticipations is in the opposite direction,

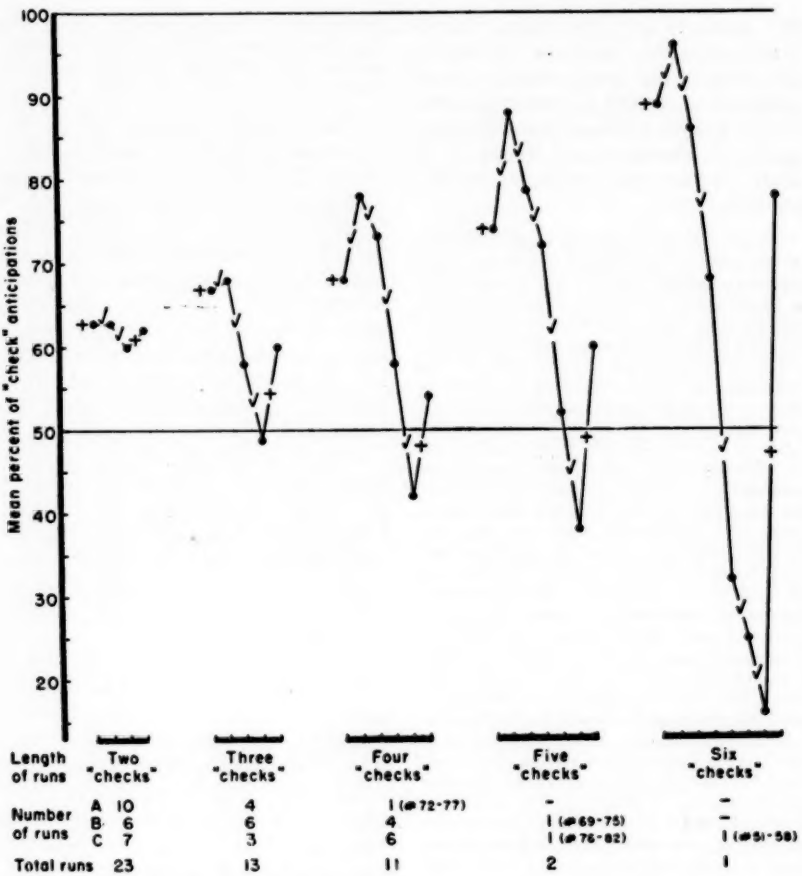

FIG. 1. Negative recency effect shown by the decline of mean "check" anticipations in incidental runs of from two to six consecutive "check" reinforcements

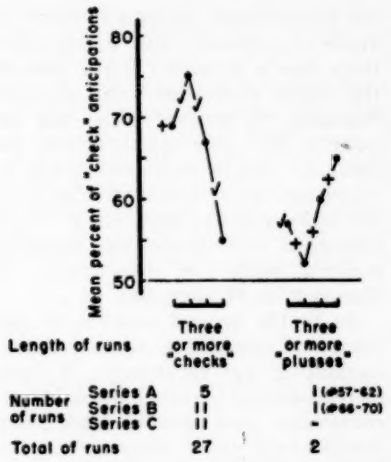

FIG. 2. Reciprocal negative recency effects in incidental runs of three or more "checks" vs. three or more "plusses"

that is, of "plusses" to come. In the single long run of six consecutive "check" reinforcements in the second half of Series C, only 18 per cent anticipate a further "check" to be presented as the "correct" guess. The "plus" presentation following each run brings expectancies back above chance or to not far from where they were before the run, however.

Continuous runs of the generally less frequent "plus" presentations are apt to be relatively short. Nonetheless, the presence of a negative recency effect for "plus" reinforcements can be observed in Fig. 2. Mean responses to the two instances of runs of over three (three and four) consecutive "plus" reinforcements yield an inverted duplicate to the means of the total of 27 runs of three or more "checks" derived from the data in Fig. 1. The right-hand curve in Fig. 2 lies somewhat lower than the left, which must be ascribed to the fact that runs of "plusses" occur in the series with lesser over-all reinforcement ratios. Since runs of unequal length are lumped together, the effect of the "plus" reinforcement immediately following each run of "checks" is not shown in the figure. The average restitution from the guess after the last "plus" reinforcement (the third or fourth, respectively) to the guess after the immediately following "check" reinforcement is −14 per cent on our scale. This recovery is of opposite sign but similar in degree to the amount of recovery for the series of comparable length in the middle part of Fig. 1.

An explanation of the negative recency effect may be attempted by pointing to the fact that conditions specific to random sequences in themselves suggest frequent alternation rather than consistency in the support of one over the other of two alternatives. Thus, in daily life a negative recency effect similar to the one observed here manifests itself in the anticipations exhibited by the average person in coin-tossing or in predicting the sex of the next in a series of children. Hence the colloquial term "gambler's fallacy" may be applied to our case. Although the logic behind the expectation of a reversal is illusory in the case of truly independent coin tossings, long runs of identical tossings are rare under the laws of chance, and over-all relative frequencies are fairly similar within more or less limited segments of the series somewhat in the manner of our experiment. Under our segmentation plan the presentation series were more reliable over time than would be the case in sequences left entirely to chance in the framework of a predetermined over-all probability. The presentations coming at the end of the segments were thus actually dependent upon those at the beginning of the segment in a negative, compensatory manner (the

more "checks" at the beginning of the segment, the fewer at the end). With full insight into our plan, the last call for each segment can be anticipated with complete accuracy. The observed negative recency effect must be seen as the result of a relatively articulated, complex learning process involving partial mastery of both the over-all trend and the more molecular superimposed segmental aspects of the presentation pattern.

The negative recency effect must be sharply distinguished from the long-range decline of the mean probability learning curve referred to in discussing the over-all trend of learning. The long-range decline is perhaps more genuinely paradoxical than the negative recency effect in that it seems to be part of an atmosphere specific to uncertainty situations, while the type of segmental learning implied in the negative recency effect may more readily be subsumed under known learning principles. Absence of a negative recency effect in the results of Brunswik and Herma (**4**) points to the lack of ecological generality (**3**) of the phenomenon.

Explicit awareness of design.—The 60 and 75 per cent groups (Series A and C), after they had completed the experiment, received the following instruction: "Now turn your protocol over and write on the back exactly what you think was happening." Responses show that explicit awareness of presentation trends was rather dim and that regularities were often assumed without actually being present. Of the *Ss* in Series A, 49 per cent indicated that there was some kind of pattern, whereas in Series C 64 per cent indicated that there was a pattern. Of those indicating that there was a pattern, in Series A about one-third (15 per cent of the total) wrote out in detail what they thought the pattern was, whereas in Series C about four-fifths of those who thought there was a pattern (50 per cent of the total) wrote out some pattern. Actually, of course, there was no pattern, but only quasi-random sequences. No attempt was made to follow up the hypothesized patterns in the actual guessing sequence of the *Ss* concerned. (Evidence on guessing patterns with two alternatives is discussed by Skinner, **10**).

As to the general problem of the degree of explicitness of probability learning in our experiment, it must be remembered that the protocol sheet containing past guesses and their positive or negative reinforcement by *E* lay open before *S*, and instructions did not forbid their inspection and study except by tacit implication. Since explicit awareness of presentation design seemed spotty at best, and frequently erroneous, despite clear indication of learning performance, it appears likely that the functionally relevant learning processes in our case took place at a level more primordial than that of intellectual verbalization.

Summary

Three randomized series in which one word held a preponderance over another at the ratio of 60, 67, and 75 per cent, respectively, were presented to different groups of college students totalling 78 who had to guess, before each new presentation, which of the two alternatives would be next.

All three proportions are above the threshold of probability discrimination under the conditions of our experiment. The trend of results resembles the probability learning curves of Brunswik and Herma (**4**) in their slow initial rise followed by a slow and "paradoxical" decline, without reaching statistical significance in this respect.

Superimposed upon, and occasionally overshadowing, the long-range positive probability learning is a short-range "negative" recency effect consisting in an increased anticipation of the opposite alternative in cases of serial accumulation of one alternative in an incidental "run."

Verbalized hypotheses concerning the sequences, if present at all, seem to have little functional relevance.

(Manuscript received April 18, 1950)

References

1. Brunswik, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, **25**, 175–197.
2. Brunswik, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255–272.
3. Brunswik, E. *Systematic and representative design of psychological experiments.* Berkeley: Univ. of California Press, 1947.
4. Brunswik, E., and Herma, H. Probability learning of perceptual cues in the establishment of a weight illusion. *J. exp. Psychol.*, 1951, **41**, 281–290.
5. Grant, D. A., and Hake, H. W. Acquisition and extinction of the Humphreys' verbal response with differing percentages of "reinforcement." *Amer. Psychologist*, 1949, **4**, 226. (Abstract.)
6. Fisher, R. A., and Yates, F. *Statistical tables for biological, agricultural and medical research.* London: Oliver and Boyd, 1938.
7. Jarvik, M. E. Probability discrimination and the gambler's fallacy in guessing. *Amer. Psychologist*, 1946, **1**, 453–454. (Abstract.)
8. Levin, M. M. Weight illusions induced by cues of low validity. *Psychol. Bull.*, 1943, **40**, 582. (Abstract.)
9. Levin, M. M. Inconsistent position cues in the establishment of perceptual illusions. Unpublished Doctor's dissertation, Univ. of California, 1946.
10. Skinner, B. F. The processes involved in the repeated guessing of alternatives. *J. exp. Psychol.*, 1942, **30**, 495–502.

THE EFFECT OF SCALE NUMBERING ON SCALE-READING ACCURACY AND SPEED[1]

BY J. L. BARBER AND W. R. GARNER

The Johns Hopkins University

The design of scales and dials for most efficient use has, since the last war, been a topic of major concern to many experimenters. Among the various factors which have been shown to affect the speed and accuracy with which dials or scales can be read, are the size and shape, the number and spacing of interval markings, and the type of indicating marker. Perhaps more difficult to demonstrate, nevertheless actively affecting scale-reading, are verbal factors such as the number system employed on the scale. This latter problem is our specific concern in this paper.

Chapanis and Leyzorek (3) and Bartlett, Reed, and Duvoison (1) have done previous work on the effect of the number system on scale-reading. In the study by Chapanis and Leyzorek it was clearly shown that the accuracy with which interpolated scale readings were made was best when the interval markers were numbered 1–2–3–4 (or 10–20–30–40). Accuracy was not significantly poorer when the interval markers were numbered 2–4–6–8, but was much poorer for the other numbering systems. The Bartlett *et al.* study used scales progressing by fives, tens, and twenties, and again the results showed the superiority of the "ten" and "twenty" scales as compared to the "five" scale. It seems fairly clear that the accuracy of interpolation between scale markers is a function of the numbering system used, all other factors being equal.

Another study of interest in this connection is one by Vernon (6). She studied the speed and accuracy of reading dials and scales to the nearest marked interval with different number systems. She found a strong effect of the number system, but as she points out, and as shown in Chapanis *et al.* (2), the more critical factor seems to be the numerical value *to which* a reading is made, not the numerical values actually appearing on the scale. Thus it was the numerical value of the smallest scale subdivision which was critical, not the numerical value of the marked divisions. The critical factor is not what the scale reader sees, but, rather, what he says. Yule (7) has previously shown that all numbers are not used equally often in reading scales, and it seems quite likely that people much prefer to use some numbers rather than others. Thus we might expect that if a scale is set up so that *S* uses numbers he likes to use, the scale will be more accurately used than if he is forced to use numbers he does not prefer. In both the Chapanis and Leyzorek study and the Bartlett *et al.* study, *S*s were allowed to use any subdivision they wished in the interpolation between marked intervals.

Purpose of experiment.—The present experiment was designed to determine more exactly the nature of the

[1] This research has been supported in part under the terms of a contract between Special Devices Center, Office of Naval Research, and The Johns Hopkins University, Contract N5-ori-166, Task Order I. This is Report No. 166-I-115, Project Designation No. NR-784-001, under that contract.

influence of scale numbering on interpolation accuracy. Scales by fives, tens, and twenties were used, but *Ss* were forced to use only certain subdivisions for their readings. They were not allowed complete freedom in making interpolations as precisely as possible. In addition, the experiment was conducted under instructions to work for speed and to work for accuracy. It was thought that the effect of the number system would be different under these two conditions.

Design of the Experiment

Accordingly, an experiment was designed using three different scales, two conditions of task instruction, and 12 *Ss*. The experiment was constructed to permit treating the data by the analysis of variance technique. The task of *S* was to determine the position of small dashes which appeared within a rectangular display. The dashes are hereafter referred to as targets. Under each instruction every *S* made 80 readings using each scale—a total of 480 readings per *S*. The presentation was arranged according to the accepted principles of experimental design. Six *Ss* first worked under an accuracy instruction and secondly under a speed instruction. With the remaining six *Ss* the order of these instructions was reversed. This procedure was adopted in an effort to minimize any confusion on the part of *S* as to which instruction he should be following at any particular time. In the analysis it results in a confounding of the subjects × instruction interaction variance, but this was an unimportant source here, and the former considerations seemed the more important.

The orders of presentation of the scales were determined by their six possible permutations. All 80 targets were read on one scale before the next scale was presented. Since each *S* used all scales under both instructions, the order of presentation under the second instruction was selected to have no sequence of scales in common with the first order given.

Targets were arrayed in eight randomized groups of ten each. These arrays were presented in random order. Each particular order of arrays was used twice in the entire experiment but never twice with the same *S*.

Apparatus

The apparatus consisted of: (1) A special projection table, the translucent top of which was illuminated from below. (2) Three photographic plates, each with one of the experimental scales imprinted on it (see Fig. 1). (3) A set of eight photographic plates bearing the targets. These are referred to as "target plates."

The three scales selected for comparison varied only in the numbers assigned to the marked intervals. These were a "five" scale (0–5–10–15–20), a "ten" scale (0–10–20–30–40), and a "twenty" scale (0–20–40–60–80). These scales were arranged along both short sides of a rectangle whose inside dimensions were 8 cm. by 13.9 cm. Three thin lines parallel to the longer sides marked the intervals. The distance, 20 mm., between markers was the same in all cases. Scale readings were made only in respect to the 8-cm. dimension.

Eighty targets were chosen and were 1 mm. apart; thus, each of 20 divisions between the interval markers was sampled four times. These 80 targets were assigned ten to a plate in such a fashion that all areas of the grid were equally sampled. The targets were short dashes approximately 2.7 mm. by .5 mm.

When assembled, the complete display consisted of a scale plate placed on top of a target plate. This assembly of two plates was placed on the top of the projection table. An opaque mask covered the area of the table not occupied by the plates. The photographic plates had been printed so as to eliminate parallax. An electric clock was used for timing the trials.

Procedure

Instructions to subjects.—Targets were to be read from left to right. On all scales *Ss* were to estimate the position to the nearest one-twentieth of the marked interval. That is, on the "five" scale they were to interpolate to $\frac{1}{4}$

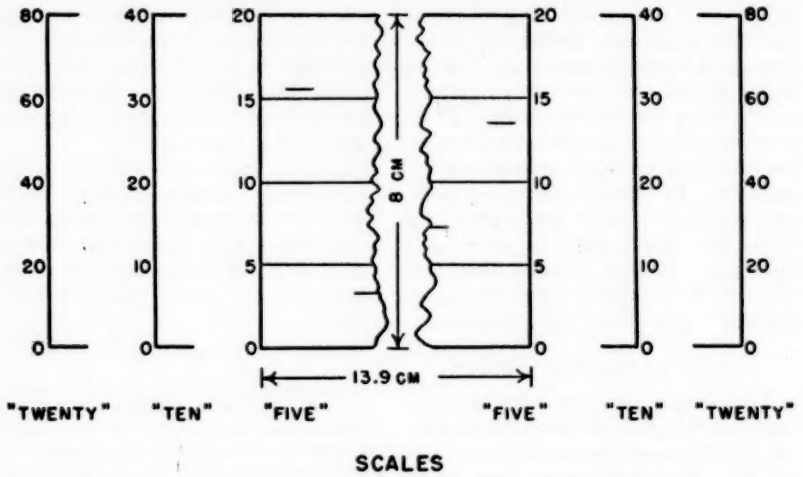

FIG. 1. Each of the three scales shown was printed along the two short sides of the rectangular display. Target positions were read with respect to these scales.

unit, on the "ten" scale to the closest ½ unit, and on the "twenty" scale to the nearest whole unit. For example, the same target would be read 7½ on the first scale, 14½ on the second and 29 on the last scale. In other words, the physical size of the unit to which readings were made was the same for all three scales.

For the accuracy instruction, *S* was told to estimate the positions as accurately as he possibly could regardless of the elapsed time. For the speed instruction, he was told not to take time to refine his judgments but to read the positions as rapidly as possible. No rigid, formal instructions were used, but an attempt was made to get across to each *S* that under accuracy instruction he should ignore time, while under the speed instruction the emphasis was primarily on how quickly the estimations could be made.

Executing the experiment.—The *S*s participated one at a time. In order to familiarize *S* with the experiment, the three scales were displayed and explained. Following that, *S* had six practice trials, or two trials with each scale. A trial consisted of estimating the position along the linear scale of each of the ten targets on one of the target plates. The experimental trials began immediately after the practice period. The *S* was given either the speed or accuracy test instruction. Then the scale to be used in the first trial was shown and the fractional unit for interpolation made clear. The *E* aligned the scale plate on top of the target plate and then darkened the display. After *S* was ready a signal was given; immediately, the display was illuminated and the clock started. The *S* called out the estimated positions to a recorder. As soon as *S* called out the position of the last of the ten targets, the clock was stopped and the display darkened. The *E* then substituted another target plate. After *S* had read all 80 targets, the procedure was repeated using a different scale. When all three scales had been tested, *S* switched to the other task instruction and again estimated the position of the 80 targets with each of the three scales.

RESULTS

Table I presents a summary of the most important data from the experiment. The time scores are simply the elapsed time between the presentation of a scale and the reporting of the position of the last of the ten targets on that scale. Thus, the times are those required to read ten targets. The error scores are average absolute errors (taken without regard to direction) expressed in percentage of the marked interval (the distance between two reference marks on the scale). An error is the discrepancy

between the true and the reported target position.

Time scores.—An analysis of variance for the time scores showed that all variances, including the subject × scale × instruction variance, are significant at beyond the .01 level when tested against the replication variance. When the triple interaction variance is used as the error estimate, only the three primary variances are significant.

In Table I, it can be seen that the "five" scale is read least rapidly and the "twenty" scale most rapidly, with the "ten" scale falling between. All differences between scales are significant in all cases except those between the "ten" and "twenty" scale under speed instructions, where the time scores are identical.

Error scores.—The analysis of variance for the error scores also showed that all variances, including the triple interactions, are significant, when tested against the replication variance. When the variances are tested against the triple interaction, however, once again only the primary variances are significant.

A glance at Table I shows that most of the variance due to the different scales comes from the fact that the "five" scale is much less accurately used than either the "ten" or "twenty" scale. There is no significant difference between the "ten" and the "twenty" scales under either instruction. All scales, however, show a sizeable increase in errors when speed rather than accuracy instructions are used.

In summary, then, the "five" scale is used less accurately and much more slowly under either speed or accuracy instructions. The only result which shows a difference between the "ten" and the "twenty" scale is that for time scores under accuracy instructions. Under this condition, the "twenty" scale is read more rapidly. Thus, this is also the only result which to any extent confirms our original hypothesis that the difference in instruction would influence the effect of the scale on speed and accuracy of reading. It is true that all of the interactions were significant, but the interaction variances constitute a small part of the total variance; the primary variances are far more important.

Confusion errors.—Another type of error, distinguishable from errors of estimation, resulted from a failure to identify the proper interval markers. Errors of this type were naturally large and were usually readily identified. Such confusion errors were tabulated separately and the estimation error corrected by adding or subtracting an amount equal to the

TABLE I

SUMMARY OF THE RESULTS

Scale	Average Time in Seconds*			Average Errors in Per Cent**		
	Instruction			Instruction		
	Accuracy	Speed	Both	Accuracy	Speed	Both
Five	80.4	40.7	60.6	3.9	5.8	4.8
Ten	75.9	34.1	55.0	2.4	3.2	2.8
Twenty	64.7	34.1	49.4	2.2	3.4	2.8

* For estimating positions of 10 targets.
** Percentages of the marked interval.

scale interval. In a few cases large errors occurred which could not be identified as simple errors in scale-reading. They may have been recording errors, and were discarded. In terms of the number of confusion errors made, the three scales did not differ more than we would expect on a chance basis.

Discussion

The question in which we were primarily interested was why one number system should be superior to another. The hypothesis we had was that the critical factor is not the number actually appearing on the scale, but, rather, the numbers used in making the reading.

Number Preferences

If we examine each of the 20 positions between the interval markers and plot for each the relative frequency with which its assigned value was used, the graphs of Figs. 2, 3, and 4 result. If all possible responses were equally represented, each position would have 5.0 per cent of the total number of reports. These graphs, however, show considerable fluctuation around this theoretical mean percentage, indicating that some numbers are used as responses much more frequently than others. Since all stimulus positions between markers were presented equally often, we can infer that the *S*s had a preference for some numbers over others. These preferences can be broken into two classes.

1. *Whole-number preferences.*— Whole numbers are used much more frequently than fractional numbers. Thus, those scales which have the fewest whole numbers to be used as responses show the greatest inequalities in the use of the values assigned to the various stimulus positions. The "five" scale shows the most serious fluctuations, and the "ten"

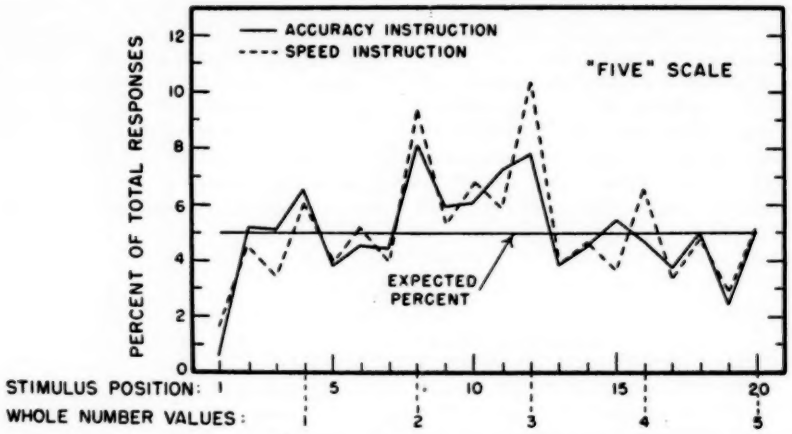

Fig. 2. Frequencies with which various stimulus positions were used as responses with the "five" scale. The actual whole-number values used with this scale are shown at the bottom of the figure. The total number of responses was approximately 950 for each of the two different instructions. If all stimulus positions had been used equally often as responses, the heavy black line would have resulted.

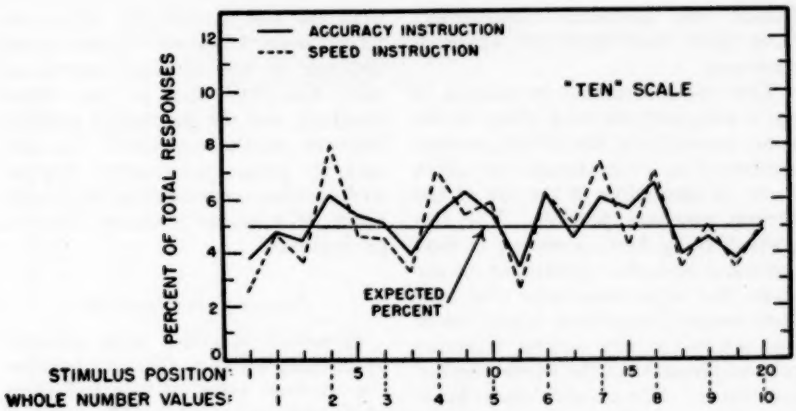

FIG. 3. Frequencies with which various stimulus positions were used as responses with the "ten" scale. Conditions are the same as for Fig. 2.

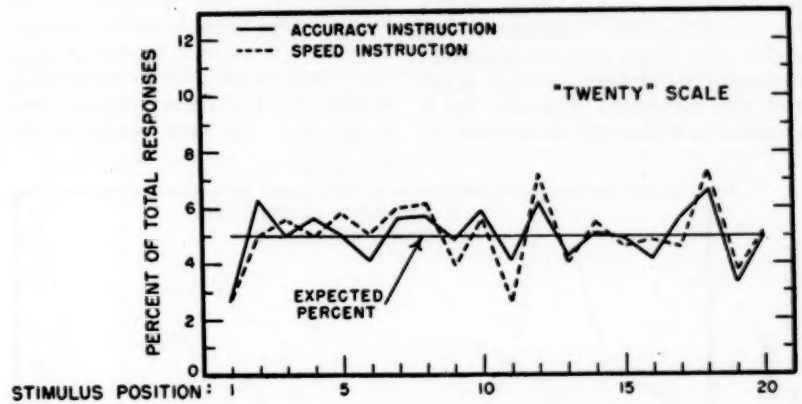

FIG. 4. Frequencies with which various stimulus positions were used as responses with the "twenty" scale. Every stimulus position was used as a whole number.

scale the next most serious fluctuations. The "twenty" scale shows some fluctuation and a tendency for even numbers to be used more frequently than odd numbers, but these fluctuations are minor compared to those found with the other two scales. In all cases, the unequal preferences are enhanced when the speed instruction is used.

2. *The central-number preference.*—Another effect which occurs primarily with the "five" scale is the tendency for the middle positions between scale markers to be used as responses more frequently than the end positions. In fact, the two curves for the "five" scale show this effect to such an extent that even the fractional numbers in the middle of the scale

(which are normally unpreferred) have more than their fair share of responses.

This central-number preference is not a number-preference effect in the same sense that the whole-number preference is, even though the effect is one of inequality in the use of the various numbers possible. The important thing here, however, is that the central-number preference occurs under the same conditions that produce serious inequalities in the use of the values for the various stimulus positions because of the whole-number preference. One possible explanation is that the number system provides to a certain extent the structuring for the perceptual problem. If the number system provides a whole number for every possible response *S* can make, then there is good agreement between the numbers *S* prefers to use and the perceptual responses he is expected to make, and we can call the situation well structured. With the "five" scale, however, *S*'s perceptual response is not in good agreement with his preference to use whole numbers, and the perceptual problem becomes poorly structured. In this case, *S*'s general tendency to respond with numbers representing the middle range of stimulus positions becomes predominant.

Average Error Functions

Figures 5, 6, and 7 were achieved by plotting for each stimulus position the average error of the judgment made of that position. The average error as computed here gives an index of the variability of the estimates around the "true" position of the target. One figure is shown for each scale, and a curve is shown for each task instruction. The outstanding feature of all of these curves is their bimodality. Of most interest is the

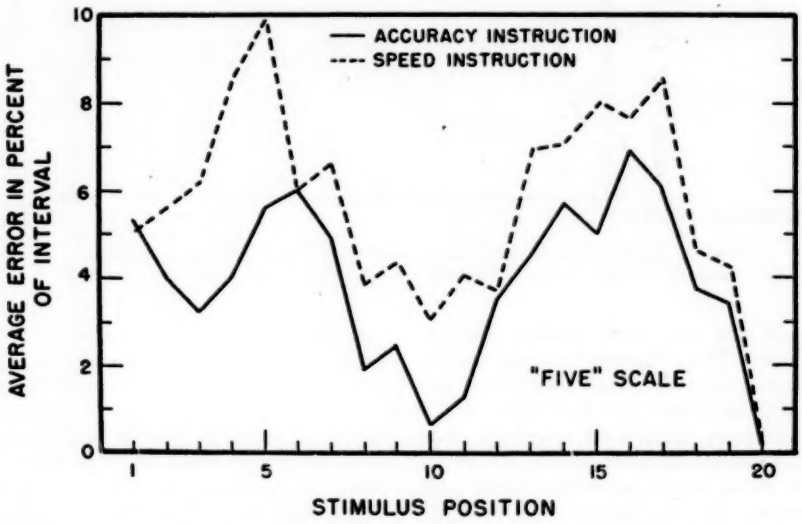

FIG. 5. The average error of interpolation as a function of the stimulus position between scale markers for the "five" scale

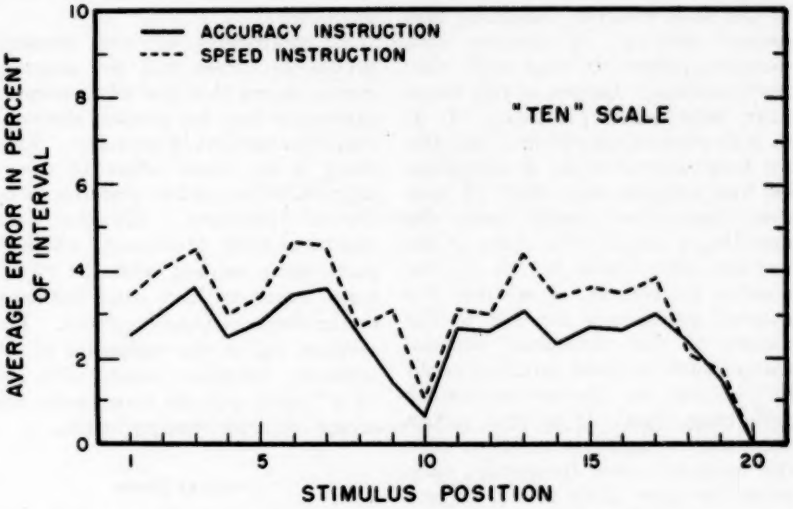

FIG. 6. The average error of interpolation as a function of the stimulus position between scale markers for the "ten" scale

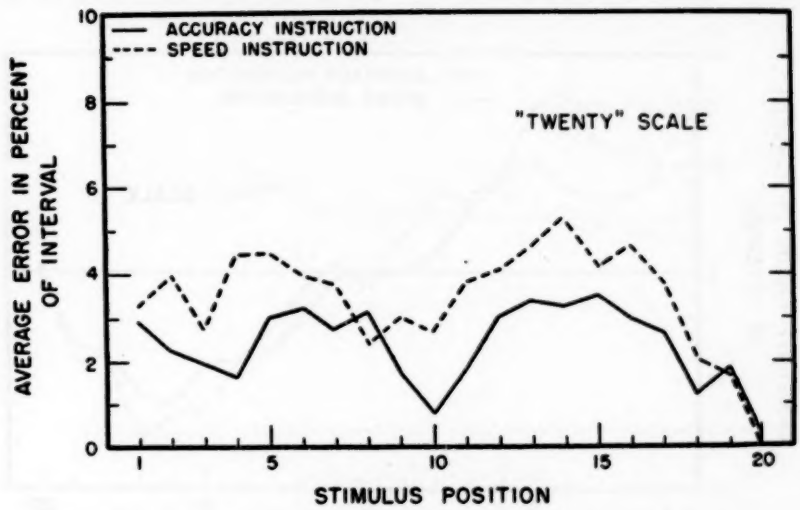

FIG. 7. The average error of interpolation as a function of the stimulus position between scale markers for the "twenty" scale

dip at positions around the midpoint of the scale interval, indicating that greater accuracy is possible with bisection judgments than with other fractionations. Curves of this shape have been noted previously (**1**, **3**) in scale-reading experiments, and also for judgments of angle of inclination (**4**) and azimuth angle (**5**). In each case there has always been the question of whether the shape of the average error curve is due to the number preferences, or whether the number preferences are due to the nature of the perceptual process. Curves such as these certainly could be produced by the central-number preference alone. If *Ss* tend to use the values assigned to positions near the midpoint more frequently, they would be more likely to call targets at these positions correctly and consequently reduce the average error. It follows that a preference for the middle positions would tend to increase the average error for other positions.

A comparison of the number-preference curves with the accuracy curves shows that the whole-number preference does not produce the same rapid fluctuations in accuracy. Thus there is no direct effect of severe differences in number preferences on over-all accuracy. However, the central-number preference, which is particularly evident with the "five" scale, seems to have some influence on the shapes of these functions. The severest dip at the midpoints of the accuracy functions occurs with the "five" scale, and the least severe dip occurs with the "twenty" scale.

Constant Errors

Constant errors as a function of target position are shown in Figs. 8, 9, and 10. The C.E. is the difference between the algebraic mean of all the

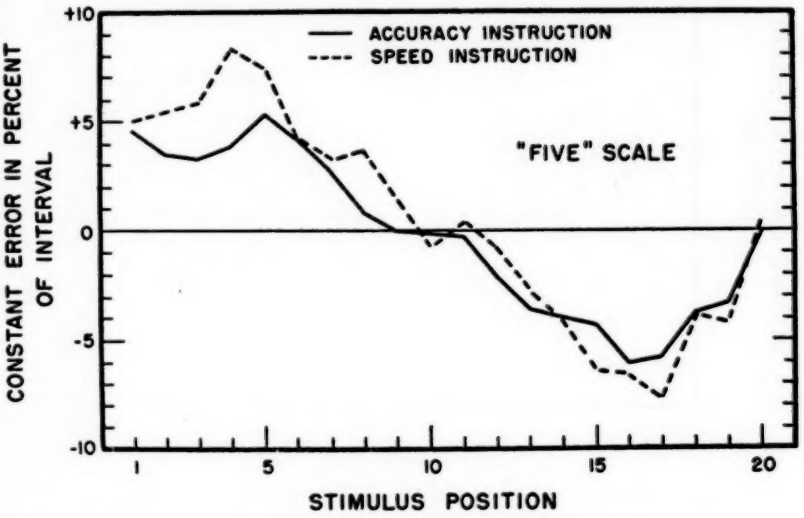

FIG. 8. The constant error of interpolation as a function of the stimulus position between scale markers for the "five" scale

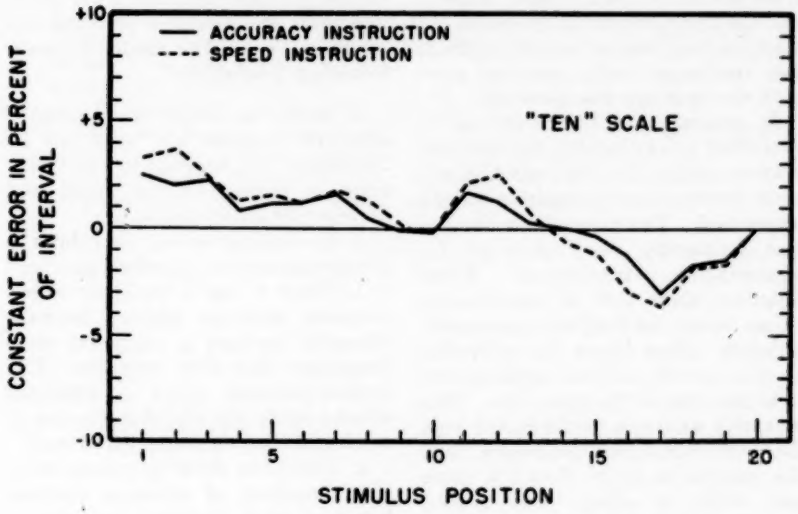

FIG. 9. The constant error of interpolation as a function of stimulus position between scale markers for the "ten" scale

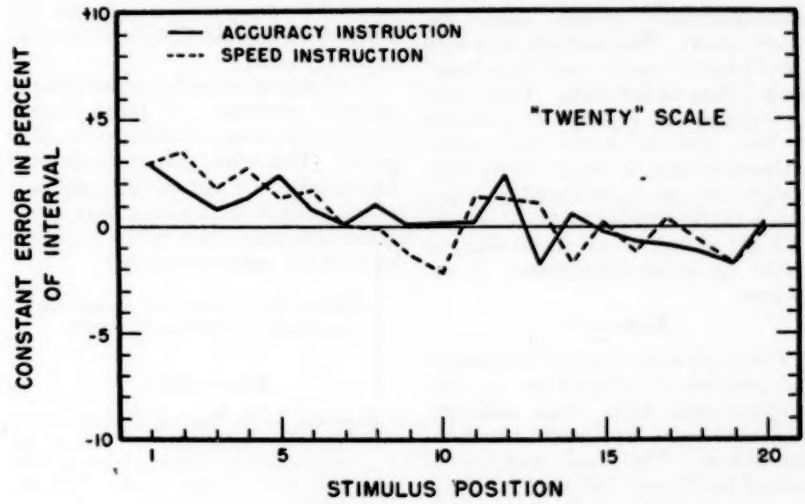

FIG. 10. The constant error of interpolation as a function of the stimulus position between scale markers for the "twenty" scale

readings for a single target position and the true value of that position. A positive error means that the estimate was too large, and a negative error that the estimate was too small.

In general, all of the curves can be described by saying that the constant error is positive for the lower half of a scale interval and is negative for the upper half. This type of function has also frequently been observed for scale-reading experiments. These functions show that *S*s overestimate values below the midpoint and underestimate values above the midpoint. In other words, all their readings tend to be too close to the midpoint. This is exactly what you would expect from the central-number preference, and the question is again: Which is cause and which is effect? The answer seems to be that at least in part, the shapes of these curves are due to the particular number system used. Notice that the most severe over- and underestimation occurs with the "five" scale. The least effect is with the "twenty" scale, with the "ten" scale falling in between. There may be a perceptual process which results in this type of over- and underestimation, but it seems likely that the greater amount of the effect is due purely to the number habits of people, and the influence of these number habits on their distribution of responses.

Summary

Twelve *S*s were required to estimate the position of targets on a rectangular grid with three different numerical scales and under two task instructions. The three scales progressed by "fives," by "tens," and by "twenties." One task instruction emphasized speed, and the other accuracy. Reading time and errors in interpolation between scale markings were the measures of speed and accuracy used. The results allow the following conclusions:

1. Scales on which the numbered intervals progress by "tens" or by "twenties" are superior to the "five" scale in respect to the speed and accuracy of interpolation.
2. In reading scales, *S*s exhibit a strong tendency to use whole numbers.
3. There is also a tendency to use stimulus positions midway between reference markers as responses more frequently than other positions. This central-tendency effect is enhanced when a scale not allowing the use of whole numbers in responses is used.
4. The curve showing average error as a function of stimulus position between scale markers exhibits a bimodality: Accuracy is greatest at the ends and the midpoint of the interval between markers. The increased accuracy at the midpoint can in part be explained by the central-tendency effect.
5. Constant error functions show a general tendency to judge target positions as being too close to the midpoint. This effect can in large part be explained by the central-tendency effect, since it occurs least with those scales showing little relative preference for the central numbers.

(Manuscript received for immediate publication October 19, 1950)

References

1. Bartlett, N. R., Reed, J. D., and Duvoison, G. Estimations of distance on polar coordinate plots as a function of the scale used. *J. gen. Psychol.*, 1949, **41**, 47–65.
2. Chapanis, A., Garner, W. R., and Morgan, C. T. *Applied experimental psychology.* New York: Wiley, 1949.

3. CHAPANIS, A., AND LEYZOREK, M. Accuracy of visual interpolation between scale markers as a function of the number assigned to the scale interval. *J. exp. Psychol.*, 1950, **40**, 655–667.
4. ROGERS, A., VOLKMANN, J., REESE, T. W., AND KAUFMAN, EDNA L. Accuracy and variability of direct estimates of bearing from large display screens. Psychophysical Research Unit, Mount Holyoke College. Memorandum Report No. 166-I-MHC 1, 1947.
5. SALTZMAN, D. C., AND GARNER, W. R. Accuracy of visual estimation of azimuth position. *J. Psychol.*, 1950, **29**, 453–467.
6. VERNON, MAGDALENE D. Scale and dial reading. Med. Res. Counc. Unit in Appl. Psychol., Cambridge Univ., England, Rep. No. APU 49, June, 1946.
7. YULE, G. W. On reading a scale. *J. roy. statist. Soc.*, 1927, **90**, 570–587.

THE ACCURACY OF COUNTING REPEATED SHORT TONES [1]

BY W. R. GARNER [2]

The Johns Hopkins University

The purpose of the present experiments was to determine the accuracy with which *Ss* can count the number of short tones in a temporal series as a function of the repetition rate of the tones and the total number of tones in the series.

Previous experimental work on the accuracy of determining the number of presented stimuli has been confined almost entirely to the field of vision.[3] In that field the work has progressed from the original experiments on span of attention to the point where a recent article (1) has distinguished counting of stimuli from the subitizing of number, a process which does not involve successive attention to each separate stimulus object. This previous work in the field of vision has been confined to numbers of stimulus objects which are spatially separate but simultaneous in time. In audition, it is not convenient to present spatially separate stimuli, but stimuli occurring successively in time are quite appropriate.

Apparatus and Procedure

In the experiments the *S* was presented a series of short 1000-cps tones, and he was required to count as accurately as he could the total number of short tones. The series varied in the total number of tones from 1 through 20, and the rates at which the tones were presented varied between 4 and 12 per sec. Any single series of tones, however, had only one rate, i.e., the tones were equally spaced in time.

Apparatus.—An oscillator which was always set at a frequency of 1000 cps provided the basic tone. This tone was put through an electronic timing unit which produced the desired tonal duration and repetition rate. A special electronic counting circuit was used to set a given number of tones. When a switch was thrown, the tones would come through, and they were automatically cut off when the desired number had been presented to the *S*. Attenuation and amplifier circuits were used to present the tones to the *S* monaurally in a PDR-8 earphone mounted in a doughnut cushion.

Procedure.—In the main experiment *S* was seated in a sound-proofed room isolated from the equipment. He could be heard by *E* through an intercommunication system. The *E* said, "Ready," and threw a switch which allowed the series of tones to come through. The *S* called out the number of tones he thought he had heard, and this number was recorded by *E*. The *E* said "Ready" before each series of tones was presented. After an average of 13 different series of tones had been presented, the rate was changed and an average of 13 series was then presented at the new rate. The 13 series presented at a single rate were randomized, and the rates were randomized. The *Ss* were run for approximately 1 hr. at a time.

Scoring system.—The main score desired was a percentage of correct counts. However, *Ss* have a strong tendency to use some numbers more often than others, and this tendency produces erratic error scores. If *S* uses the number *10* five times as often as he uses the

[1] This research was done under the terms of a contract between Special Devices Center, Office of Naval Research, and The Johns Hopkins University, Contract N5-ori-166, Task Order I. This is report No. 166-I-122, Project Designation No. NR 784-001, under that contract.

[2] The author gratefully acknowledges the assistance of Mrs. Beverly Richards and Miss Jeanne Jordan in the collection and analysis of the data reported here.

[3] Since this study was completed and the report originally written, a paper by Taubman (2) has appeared which very closely duplicates the work reported here. Although the present study used a wider range of numbers and a few more control conditions, the general results and conclusions are almost identical to those reported by Taubman. Since the present report tends simply to be a confirmation of Taubman's results, the analysis and discussion has been kept to a minimum.

number *11*, then he will be correct when the presented number is *10* five times as often as when the presented number is *11*.

In order to offset this effect, we used a final score which is the geometric mean of two other scores. One of these scores is the number of correct responses divided by the number of stimuli presented at a particular number. The other score is the number of correct responses divided by the number of times that number was used as a response. For example, if the number *10* was presented 10 times, and was called correctly 5 times, the percentage would be 50. If he used that number 20 times, the percentage of correct uses would be 25. The geometric mean of these two (35.6) would be the final score actually used.

Results

Effect of intensity and duration on accuracy.—In the first experiment two tone durations and two tone intensities were used. The two intensities were 55 and 95 db. SPL. The two durations were 5 and 40 msec. The 5-msec. tone is heard as a click, while the 40-msec. tone is heard as a tone with definite pitch. Each number of tones in the series at each repetition was presented a total of 20 times to each *S* for each of the four conditions of intensities and durations. The data as a whole show no difference in counting accuracy due to either duration or intensity of each tone.

Effect of repetition rate and number of tones in series.—Figure 1 shows accuracy of counting as a function of the number of tones in the series and the repetition rate. Five *S*s were used to obtain these data, which are the combined data for all tone durations and intensities. The results clearly show that accuracy of counting is a joint function of both repetition rate and the number of tones in the series. At a repetition rate of four

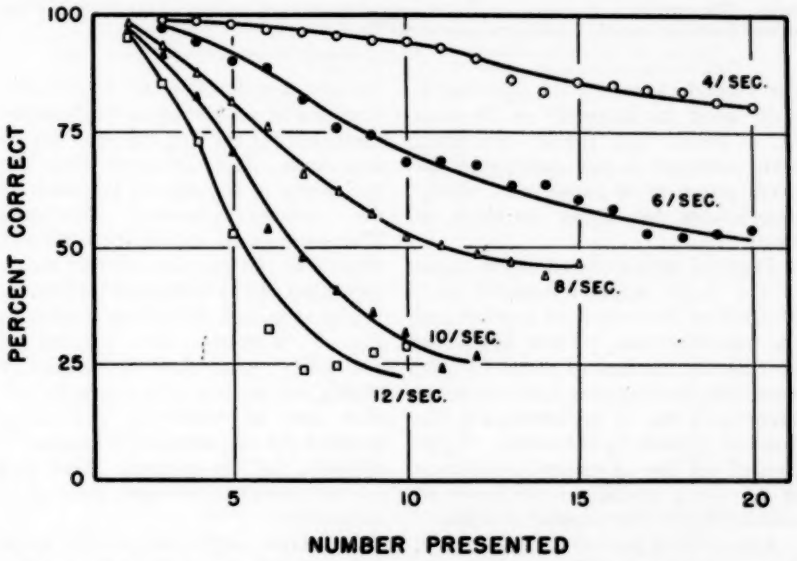

Fig. 1. The accuracy of counting repeated short tones as a function of the number of tones in a series for different repetition rates. These data are combined from 20 observations for each point from each of five *S*s under four different conditions.

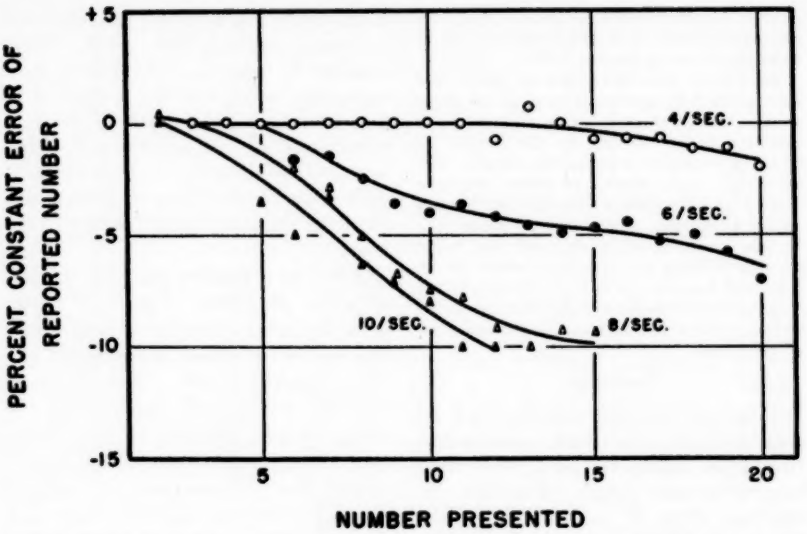

FIG. 2. The constant error of the mean reported number as a function of the number of tones in a series for different repetition rates. The constant error was computed as a percentage of the number presented, and a negative error means that *S* underestimated the number of tones in the series. The curve for a repetition rate of 12 per second was omitted because it is practically identical to that for 10 per second. Conditions same as for Fig. 1.

per second, accuracy of counting is quite good for as many as 20 tones in a series. At higher repetition rates accuracy is still quite good for small numbers of tones, but rapidly deteriorates for larger numbers of tones.

Figure 2 shows the constant errors of the mean reported number as a function of the presented number and the repetition rate. These data show that as the number of tones and the repetition rate increase (and accuracy decreases), the *S*s underestimate the number of tones in the series. Thus, part of the loss of counting accuracy is due to a consistent tendency to underestimate the number of tones.

A far greater part of the inaccuracy, however, is due to the variability in reported number as the presented number increases. Figure 3 shows the standard deviation of the reported numbers as a function of the number presented for the five different repetition rates. It is apparent that the variability of the reports increases as the number presented increases. This increase in variability is closely related to the per cent correct score, regardless of the repetition frequency.

The standard deviations shown in Fig. 3 represent two sources of variability: that due to variability within the reports of a single *S*, and that due to consistent differences between the mean reported number of different *S*s. In general, these two sources contribute an equal amount of variability.

Individual differences.—The large differences between *S*s in mean reported number, and the differences in per cent correct reports, make it

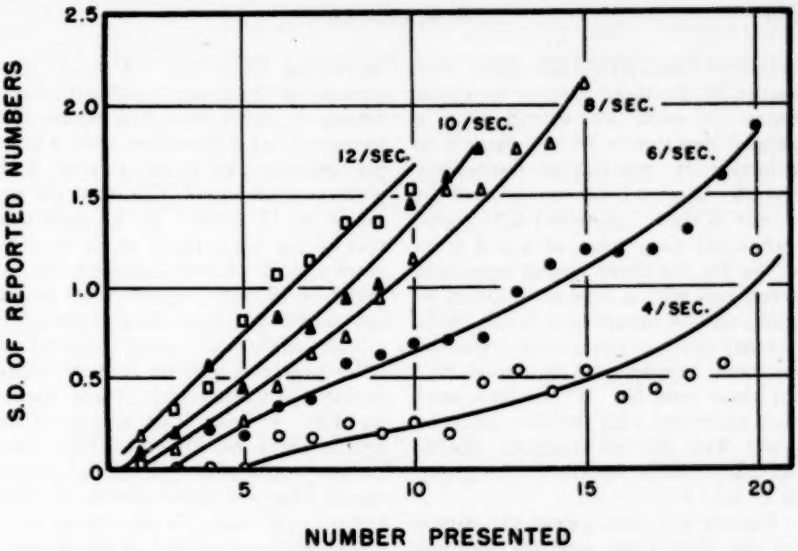

FIG. 3. The standard deviations of the reported numbers as a function of the number of tones in a series for different repetition rates. The standard deviations include both variability within the reports of single *Ss* and variability between the mean reported number for different *Ss*. Conditions same as for Fig. 1.

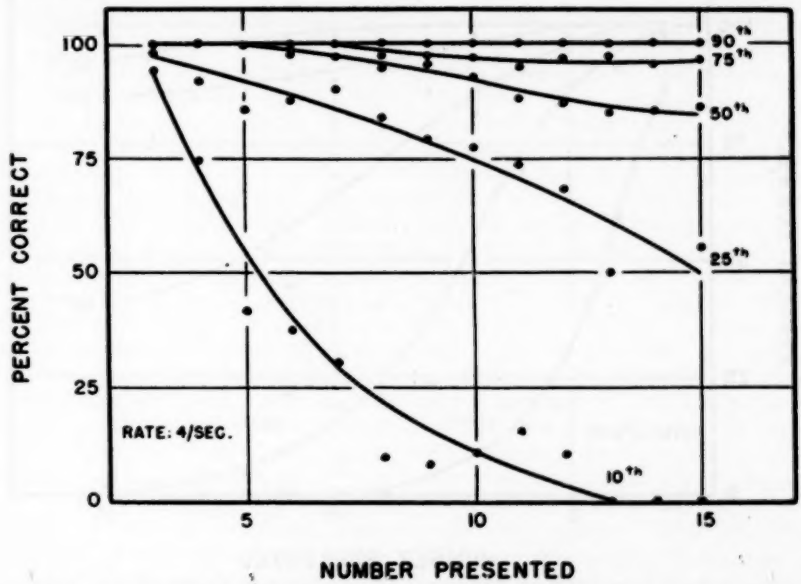

FIG. 4. The accuracy of counting repeated short tones as a function of the number of tones in a series for different group centiles at a repetition rate of four per second. These data are based on 20 observations for each point from each of 25 *Ss* under one condition.

apparent that individuals differ very markedly in their ability to count tones—to such an extent that a general description of the pattern of behavior is practically impossible. In order to get a better estimate of the extent of these individual differences, additional data were obtained from 20 *Ss* for the three slower repetition rates, and with a tone duration of 40 msec. and an intensity of 95 db. SPL. Twenty series of tones were presented for each number and repetition rate for these new *Ss*. These data were then combined with the data from the other five *Ss* to compute centile accuracy scores for the entire group of 25 *Ss*.

Figures 4, 5, and 6 show the curves for the three rates used for different centiles of the group. It is immediately apparent that there are large individual differences. A small percentage of the group could not count accurately more than four successive tones even at a repetition rate of four per second. An equal number had perfect accuracy at this rate for as many as 15 tones. At a repetition rate of six per second, some *Ss* still gave almost perfect accuracy for as many as 15 tones, while others could not even count four tones accurately.

The extent of these individual differences can best be seen by comparing the twenty-fifth centile curve on Fig. 4 with the seventy-fifth centile curve on Fig. 6. These two curves are almost identical, which means that the best quarter of the group could count as accurately at a rate of eight per second as the poorest quarter could at a rate of four per second. The original group of five

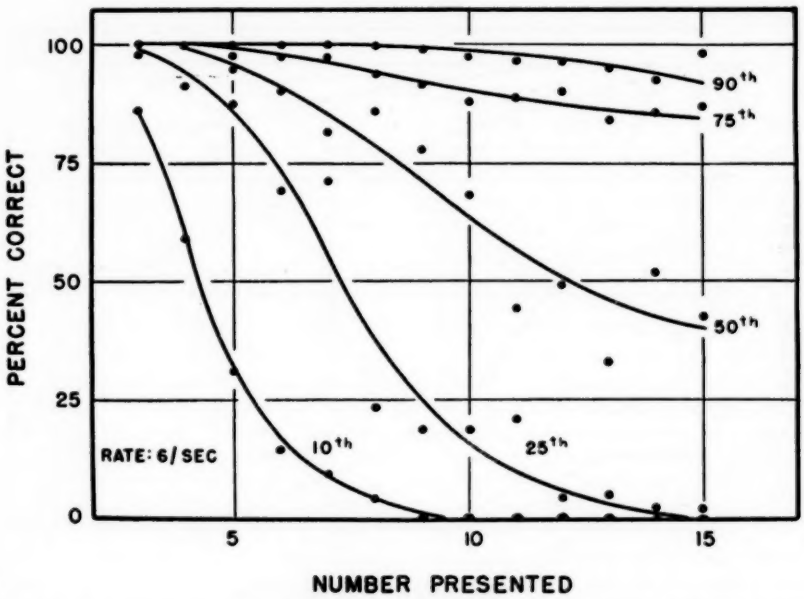

FIG. 5. The accuracy of counting repeated short tones as a function of the number of tones in a series for different group centiles at a repetition rate of six per second. Conditions same as for Fig. 4.

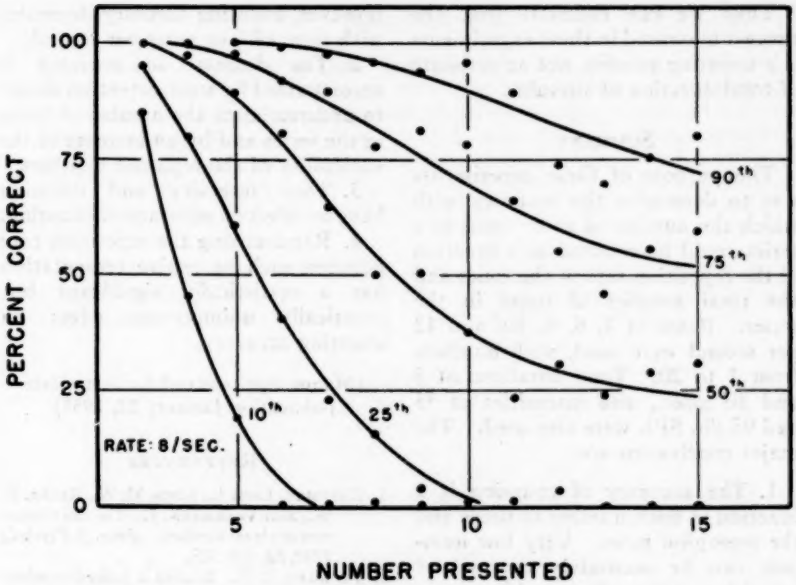

FIG. 6. The accuracy of counting repeated short tones as a function of the number of tones in a series for different group centiles at a repetition rate of eight per second. Conditions same as for Fig. 4.

Ss, from whom the most data were obtained, were not very representative of the larger group. Three of the five were in the top five from the total group.

Effect of randomizing rate of presentation.—Throughout the experiments reported so far, *S* heard several series of tones in a sequence at the same repetition rate. It seemed quite possible that after a few of these he could become more accurate in his counts, since if he could make an accurate estimate of the total elapsed time, he could arrive at the correct count. Because of equipment limitations, it was impractical to change the repetition rate between each successive series of numbers. However, in order to determine the importance of this factor, ten additional *Ss* were used in a short check experiment. Rates of 4, 6, and 8 per sec. were used, with numbers from 1 through 15. These stimuli were put on a tape recorder and played back to the *Ss* for a total of ten counts for each number of tones. Part of the time the rate was allowed to remain the same for 15 series of tones in a sequence. The rest of the time the rate was changed in a random order between each successive series of tones. If we combine all rates, numbers, and *Ss*, the total per cent correct reports for the randomized rate condition was 42.1; for the nonrandomized rate condition, 44.5. This difference between the two conditions is statistically significant, but of no practical importance. The difference due to the randomizing of the rate is insignificant compared to the differences due to rate and number of tones.

Thus we can conclude that the process measured in these experiments is a counting process, not an estimate of total duration of stimulus.

Summary

The purpose of these experiments was to determine the accuracy with which the number of short tones in a series could be counted as a function of the repetition rate of the tones and the total number of tones in the series. Rates of 4, 6, 8, 10, and 12 per second were used, with numbers from 1 to 20. Tone durations of 5 and 40 msec., and intensities of 55 and 95 db. SPL were also used. The major conclusions are:

1. The accuracy of counting is a function of both number of tones and the repetition rates. Very low numbers can be accurately counted at repetition rates as high as 12 per second. At numbers above five or six, however, counting accuracy decreases with rates of 6 or more per second.

2. The decrease in accuracy is accompanied by a consistent tendency to underestimate the number of tones in the series and by an increase in the variability of the reported numbers.

3. Tone intensity and duration have no effect on accuracy of counting.

4. Randomizing the repetition rate between each successive presentation has a statistically significant but practically unimportant effect on counting accuracy.

(Manuscript received for immediate publication January 22, 1951)

References

1. Kaufman, Edna L., Lord, M. W., Reese, T. W., and Volkmann, J. The discrimination of visual number. *Amer. J. Psychol.*, 1949, **62**, 498–525.
2. Taubman, R. E. Studies in judged number: I. The judgment of auditory number. *J. gen. Psychol.*, 1950, **43**, 167–194.

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue, N.W.
Washington 5, D. C.

The American Psychological Association, founded in 1892 and incorporated in 1925, is the major psychological organization in the United States. With approximately 8,500 members and 2,000 affiliates, it includes most of the qualified psychologists in the country. The purpose of the APA is to advance psychology as a science, as a profession, and as a means of promoting human welfare. It attempts to further these objectives by holding annual meetings, publishing psychological journals, conducting a personnel placement service for psychologists, and working toward improved standards for psychological training and service.

In order to give recognition to the specialized interests of different psychologists, the APA includes seventeen Divisions. Any person, after becoming a member of the APA, may apply for membership in as many Divisions as he wishes. The Divisions are:

Division of General Psychology
Division on the Teaching of Psychology
Division of Experimental Psychology
Division on Evaluation and Measurement
Division on Childhood and Adolescence
Division of Personality and Social Psychology
The Society for the Psychological Study of Social Issues—a Division of the American Psychological Association
Division on Esthetics
Division of Clinical and Abnormal Psychology
Division of Consulting Psychology
Division of Industrial and Business Psychology
Division of Educational Psychology
Division of School Psychologists
Division of Counseling and Guidance
Division of Psychologists in Public Service
Division of Military Psychology
Division on Maturity and Old Age

Each division has its own officers. Each meets annually at the time and place of the APA meeting. Each has its own membership requirements, which in some cases are higher or more specialized than the requirements for election to the APA.

The annual meetings of the APA are held in September. An attempt is made to choose meeting sites in such a way that attendance will be convenient for members and their guests in different sections of the country at different times. Thus, the 1951 meeting will be held from August 31 through September 5 in Chicago, and the 1952 meeting in Washington, D. C. These meetings provide the members with an opportunity to present or hear reports of psychological research and lectures on topics of psychological interest, and to participate in formal and informal discussion groups. The research reports, lectures, and discussion groups cover varied topics of current interest within the field of psychology.

The chief governing body of the APA is the Council of Representatives. It includes representatives from each of the Divisions. A Board of Directors, composed of the four officers of the APA (President, President-elect, Recording Secretary, and Treasurer) and six Council members elected by the Council, is the administrative agent of the Council and exercises general supervision over the affairs of the Association.

The APA maintains an office in Washington which acts as a coordinating center for all APA activities. Journal subscriptions, membership applications, personnel placement requests, and general APA business procedures are handled in this office.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The issues listed below are available for sale

Year	Volume	Available Numbers						Price per Number	Price per Volume
1916	1	1	2	3	4	5	6	$1.25	$7.00
1917	2	1	2	–	4	5	–	$1.25	$5.00
1920	3	–	2	3	4	5	6	$1.25	$6.25
1921	4	1	2	3	4	5	6	$1.25	$7.00
1922	5	–	–	3	4	5	6	$1.25	$5.00
1923	6	–	–	–	–	–	–	——	——
1924	7	1	2	–	–	5	–	$1.25	$3.75
1925	8	1	2	3	4	–	6	$1.25	$6.25
1926	9	–	2	3	–	5	6	$1.25	$5.00
1927	10	–	–	3	4	5	6	$1.25	$5.00
1928	11	1	2	3	4	5	6	$1.25	$7.00
1929	12	1	2	3	4	5	6	$1.25	$7.00
1930	13	1	2	3	4	5	6	$1.25	$7.00
1931	14	1	2	3	4	5	6	$1.25	$7.00
1932	15	1	2	3	4	5	6	$1.25	$7.00
1933	16	1	2	3	4	5	6	$1.25	$7.00
1934	17	1	2	3	4	5	6	$1.25	$7.00
1935	18	1	2	3	4	5	6	$1.25	$7.00
1936	19	1	2	3	4	5	6	$1.25	$7.00
1937	20	1	2	3	4	5	6	$1.25	$7.00
	21	1	2	3	4	5	6	$1.25	$7.00
1938	22	1	2	3	4	5	6	$1.25	$7.00
	23	1	2	3	4	5	6	$1.25	$7.00
1939	24	1	2	3	4	5	6	$1.25	$7.00
	25	1	2	3	4	5	6	$1.25	$7.00
1940	26	1	2	3	4	5	6	$1.25	$7.00
	27	1	2	3	4	5	6	$1.25	$7.00
1941	28	1	2	3	4	5	6	$1.25	$7.00
	29	1	2	3	4	5	6	$1.25	$7.00
1942	30	1	2	3	4	5	6	$1.25	$7.00
	31	1	2	3	4	5	6	$1.25	$7.00
1943	32	1	2	3	4	5	6	$1.25	$7.00
	33	1	2	3	4	5	6	$1.25	$7.00
1944	34	1	2	3	4	5	6	$1.25	$7.00
1945	35	1	2	3	4	5	6	$1.25	$7.00
1946	36	1	2	3	4	5	6	$1.25	$7.00
1947	37	1	2	3	4	5	6	$1.25	$7.00
1948	38	1	2	3	4	5	6	$1.25	$7.00
1949	39	1	2	3	4	5	6	$1.25	$7.00
1950	40	1	2	3	4	5	6	$1.25	$7.00
1951	41								
	42	By subscription, $14.00 (2 vols.)						$1.25	——
		List price, Volumes 1 through 40							$260.25
		30% discount							78.08
		Net price, Volumes 1 through 40							$182.17

Information about the Journal of Experimental Psychology: as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year; from 1944 to 1950, in one volume per year. Beginning in 1951 there are two volumes a year.

Information about prices: when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Address orders and subscriptions to

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue N.W. Washington 5, D. C.